



UNIVERSIDADE TÉCNICA DE LISBOA
FACULDADE DE MOTRICIDADE HUMANA



Tradução e adaptação da *PEDro Scale* para a cultura portuguesa: um instrumento de avaliação de ensaios clínicos em Fisioterapia

Dissertação elaborada com vista à obtenção do Grau de Mestre em
Ciências da Fisioterapia

Orientador: Professor Doutor Jan Maria Hendrick Cabri

Júri:

Presidente:

Professor Doutor Jan Maria Hendrick Cabri

Vogais:

Professora Doutora Maria Margarida Marques Rebelo Espanha

Professor Doutor João António Neves Gil

Autora:

Cláudia Maria Lima Costa

2011



UNIVERSIDADE TÉCNICA DE LISBOA
FACULDADE DE MOTRICIDADE HUMANA



Tradução e adaptação da *PEDro Scale* para a cultura portuguesa: um instrumento de avaliação de ensaios clínicos em Fisioterapia

Dissertação elaborada com vista à obtenção do Grau de Mestre em
Ciências da Fisioterapia

Orientador: Professor Doutor Jan Maria Hendrick Cabri

Júri:

Presidente:

Professor Doutor Jan Maria Hendrick Cabri

Vogais:

Professora Doutora Maria Margarida Marques Rebelo Espanha

Professor Doutor João António Neves Gil

Autora:

Cláudia Maria Lima Costa

2011

"A ignorância afirma ou nega veementemente, a ciência duvida."

(Voltaire)

Agradecimentos

Não esquecer:

Serviço de Medicina Física e Reabilitação do Hospital Garcia de Orta - Dra. Índia Remédios, Ft. José Eduardo Ressurreição, Ft. Rosário Ferreira e todos os demais colegas,

Por todo o apoio.

Mestre Ana Sequeira, Mestre Ana Cristina Inácio, Prof. Doutora Ângela Pereira, Mestre João Paulo Sousa, Prof. Doutor Raul Oliveira e Mestre Sofia Pinto,

Por toda a paciência e disponibilidade.

Escolas Superiores de Saúde: ESSEM – Prof. Doutora Ângela Pereira; ESSA - Mestre Patrícia Almeida; ESTEsL - Mestre Isabel Coutinho; ESSUA – Prof. Doutor Rui Costa, e ESTEsCoimbra - Prof. Doutor João Gil,

Por toda a atenção e acessibilidade.

Bibliotecário Luís Cordeiro do HGO, Prof. Ana Carita da FMH e Mestre Mafalda Mendes

Por toda a ajuda e esclarecimentos.

Meus amigos e colegas de mestrado,

Pelo forte incentivo e amizade.

Meu orientador Prof. Doutor Jan Maria Cabri,

Por acreditar em mim e chamar à atenção quando é preciso.

“Meus doentes”,

Por querer sempre fazer mais por vós.

Minha família e a minha cara-metade,

Por todos os momentos que lhes privei da minha presença e pela compreensão e carinho que sempre tiveram.

Acredito que não teria conseguido alcançar este feito se não fosse por vós.

Um grande BEM-HAJA a todos!

Resumo

Introdução: A prática baseada na evidência tem-se tornado uma das preocupações centrais da investigação clínica em qualquer profissão de saúde que procure melhorar a sua credibilidade científica. Torna-se assim necessária a existência de um instrumento de medida que permita analisar e avaliar de forma crítica, objectiva e criteriosa o rigor metodológico e científico dos estudos publicados. A *PEDro Scale*, desenvolvida por Moseley et al tem como objectivo qualificar ensaios clínicos randomizados e ensaios clínicos controlados a partir de critérios objectivos e uniformes. O facto de não estar traduzida para a língua portuguesa constitui uma limitação para a utilização correcta do instrumento por esta população.

Objectivo: Traduzir e adaptar a *PEDro Scale* para a cultura portuguesa.

Metodologia: Foi utilizado o método de tradução/retrotradução. A 1ª versão foi analisada por um painel de 6 peritos (4 ♀ e 2 ♂ - fisioterapeutas e docentes do ensino superior). A análise foi executada através do processo Delphi informatizado, para obter a equivalência linguística e cultural. Esta versão foi submetida a um pré-teste com uma amostra de 20 sujeitos (12 ♀ e 8 ♂ - fisioterapeutas), para preenchimento da escala e respectivo formulário de compreensão.

Resultados e Discussão: No processo Delphi realizaram-se quatro rondas para poder obter a equivalência nas duas culturas. No pré-teste o nível de concordância foi 75 % e o resultado do formulário de compreensão foi igual a 4 numa escala de 0 a 5, para todos os itens. Estes resultados asseguram a boa tradução e adaptação para a cultura portuguesa.

Conclusões: A *PEDro Scale* demonstrou ter equivalência linguística à Escala de PEDro nas duas culturas assim como validade de face para ser usado em Portugal. Necessita porém de futuros estudos para averiguar as restantes características psicométricas.

Palavras-chave: Prática baseada na evidência; Avaliação de ensaios clínicos; *PEDro Scale*; Fisioterapia, Investigação

Abstract

Introduction: Evidence-based practice has become one of the central concerns of clinical investigation in any health profession that seeks better scientific credibility. Thus becomes necessary a measuring tool which allows analysing and evaluating in a critical, objective and wise way the methodologic and scientific rigor of the published studies. The PEDro Scale, developed by Moseley et al, has as its objective qualify randomized clinical trials and clinical controlled trials by using objective and uniform criteria. The fact that it is not translated to the Portuguese language constitutes a limitation for the correct use of the tool by this population.

Objective: To translate and adapt the PEDro Scale for the portuguese culture.

Methods: It was used the method of translation/ back translation sequential form. The 1st version analysis was made through the Delphi computerized process, by a 6 expert painel (4 ♀ and 2 ♂ -physiotherapists and superior education teachers), to achieve the linguistic and cultural equivalence. This version was submitted to a pretest with a sample of 20 subjects (12 ♀ and 8 ♂ - physiotherapists), for the scale filling and respective understanding form.

Results and Discussion: The consensus and equivalence for both cultures was obtained after 4 rounds on the Delphi process. The pretest level of agreement was 75% and the understanding form result was superior to 4 on a scale from 0 to 5, for all items. These results ensure the translation and adaptation for the portuguese culture.

Conclusions: PEDro Scale has demonstrated its linguistic equivalence to Escala de PEDro in both cultures as well as the face validity to be used in Portugal. However, it needs future studies to ascertain the remaining psychometric properties.

Keywords: Evidence-base practice; Clinical trials Assessment; PEDro Scale; Physiotherapy; Investigation.

Índice geral

RESUMO	IX
ABSTRACT	XI
ÍNDICE GERAL	XIII
ÍNDICE DE TABELAS	XVII
ÍNDICE DE FIGURAS	XIX
LISTA DE ABREVIATURAS.....	XXI
INTRODUÇÃO	3
REVISÃO DA LITERATURA	7
2.1 INVESTIGAÇÃO E PRÁTICA BASEADA NA EVIDÊNCIA	7
2.1.1 Níveis de evidência e graus de recomendação.....	8
2.1.2 Ensaios clínicos	10
2.1.2.1 Definição.....	11
2.1.2.1 Metodologia	11
2.1.2.3 Vantagens e Desvantagens.....	13
2.1.2.4 Regras para exposição de Relatório - CONSORT.....	14
2.1.2.2 Contribuição para a PBE.....	15
2.1.3 Revisões Sistemáticas.....	15
2.1.3.1 Definição.....	15
2.1.3.2 Metodologias Actuais	16
2.1.3.2 Regras para exposição de relatório de RS - QUORUM - PRISMA	16
2.1.3.3 Contribuição para a PBE.....	17
2.1.4 Instrumentos de Medida de Ensaios Clínicos	17
2.1.4.1 Descrição e relevância	17
2.1.4.2 Aplicações.....	18
2.1.4.3 Exemplos	18
2.2 PHYSIOTHERAPY EVIDENCE DATABASE	19
2.3 PEDRO SCALE.....	22
2.3.1 Descrição do Instrumento.....	22
2.3.2 Fundamentos e objectivos.....	23
2.3.3 Escalas, conceitos e pontuações	23
2.3.4 Administração e aceitabilidade	24
2.3.5 Fiabilidade e validade	24
2.3.6 Campos de aplicação.....	24
2.4 TRADUÇÃO E ADAPTAÇÃO CULTURAL DE UM INSTRUMENTO.....	25
2.4.1 Descrição e relevância	25
2.4.2 Etapas descritas.....	25
2.4.1.1 Preparação.....	25
2.4.1.2 Tradução	26
2.4.1.3 Reconciliação - Processo Delphi.....	26
2.4.1.4 Retrotradução.....	26

2.4.1.5 Relatório retrotradução vs versão original	26
2.4.1.6 Harmonização	26
2.4.1.7 Pré-teste	26
2.4.1.8 Análise resultados pré-teste.....	27
2.4.1.9 Revisão final	27
2.4.1.10 Versão final e Relatório	27
2.5 VALIDAÇÃO DE UM INSTRUMENTO	27
2.5.1 Descrição e relevância	27
2.6 PROCESSO DELPHI	28
2.6.1 Descrição.....	28
2.6.2 Objectivo.....	30
2.6.3 Método	30
2.6.4 Procedimento.....	31
2.6.5 Vantagens	31
2.6.6 Desvantagens.....	32
METODOLOGIA	35
3.1 DESENHO DE ESTUDO	35
3.2 OBJECTIVO DE ESTUDO	35
3.3 PROCEDIMENTOS DA TRADUÇÃO E ADAPTAÇÃO CULTURAL.....	35
3.3.1 Preparação	36
3.3.2 Tradução.....	36
3.3.3 Reconciliação – Processo Delphi	36
3.3.3.1 Selecção do painel de peritos	36
3.3.3.2 Recolha de dados peritos.....	36
3.3.3.3 Critérios de avaliação.....	37
3.3.4 Retrotradução.....	38
3.3.5 Retrotradução vs original.....	38
3.3.6 Pré-teste.....	39
3.3.6.1 Selecção dos participantes no estudo	39
3.3.6.2 Protocolo.....	39
3.3.6.3 Recolha de dados	40
3.3.7 Análise resultados Pré-teste	40
3.3.8 Revisão final	40
3.3.9 Versão Final e Relatório.....	40
3.3.10 Validação.....	41
RESULTADOS.....	45
4.1 TRADUÇÃO E RETROTRADUÇÃO	45
4.2 ADAPTAÇÃO CULTURAL.....	45
4.2.1 Processo Delphi.....	45
4.2.1.1. Caracterização do painel de peritos.....	45
4.2.1.2 Ronda 1	45
4.2.1.3 Ronda 2.....	51
4.2.1.4 Ronda 3.....	54
4.2.1.5 Ronda 4.....	54
4.2.2 Pré-teste.....	55
4.2.2.1 Caracterização da amostra do pré-teste	55

4.2.2.2 Escala de PEDro - Concordância artigo	56
4.2.2.3 Formulário de compreensão	58
4.2.2.4 Tempo de preenchimento	60
DISCUSSÃO E CONCLUSÃO	63
5.1 PROCESSO DELPHI	63
5.2 PRÉ-TESTE	63
5.3 CONCLUSÕES	64
CONSIDERAÇÕES FINAIS.....	67
6.1 LIMITAÇÕES	67
6.2 FUTUROS PROJECTOS.....	67
REFERÊNCIAS BIBLIOGRÁFICAS	71
APÊNDICES	
APÊNDICE I – PEDIDO DE AUTORIZAÇÃO A AUTORES DA PEDRO SCALE	
APÊNDICE II – RELATÓRIO ENVIADO AOS AUTORES “DRAFT PROTOCOL”	
APÊNDICE III – TRADUÇÃO PEDRO SCALE DE INGLÊS PARA PORTUGUÊS 1	
APÊNDICE IV – RONDAS PROCESSO DELPHI	
APÊNDICE V – TRADUÇÃO PEDRO SCALE DE INGLÊS PARA PORTUGUÊS 2	
APÊNDICE VI – RETRO-TRADUÇÃO	
APÊNDICE VII – CONSENTIMENTO INFORMADO PRÉ-TESTE	
APÊNDICE VIII – FORMULÁRIO PRÉ-TESTE	
APÊNDICE IX – ESCALA PEDRO (VERSÃO PORTUGUESA)	
APÊNDICE X – RELATÓRIO FINAL	
ANEXOS	
ANEXO I – PEDRO SCALE (VERSÃO ORIGINAL)	
ANEXO II – AUTORIZAÇÃO AUTORES PEDRO SCALE	
ANEXO III – RELATÓRIO RETROTRADUÇÃO VS VERSÃO ORIGINAL PEDRO SCALE	

Índice de tabelas

TABELA 2.1 – HIERARQUIA DE EVIDÊNCIA CLÍNICA.....	9
TABELA 2.2 – NÍVEIS DE EVIDÊNCIA I	9
TABELA 2.3 – NÍVEIS DE EVIDÊNCIA II.....	10
TABELA 2.4 – GRAU DE RECOMENDAÇÃO DE EVIDÊNCIA	10
TABELA 2.5 – ESCALAS DE AVALIAÇÃO DE ECRS.....	18
TABELA 2.6 – <i>PEDRO SCALE</i> ORIGINAL	23
TABELA 2.7 – TIPOS DE FIABILIDADE E VALIDADE	28
TABELA 3.1 – EQUIVALÊNCIA DOS TERMOS DIFERENTES DA RETROTRADUÇÃO.....	38
TABELA 4.1 – REFORMULAÇÃO QUESTÃO 4 – RONDA 1	46
TABELA 4.2 – REFORMULAÇÃO QUESTÃO 9 – RONDA 1	46
TABELA 4.3 – REFORMULAÇÃO QUESTÃO 11 – RONDA 1	47
TABELA 4.4 – REFORMULAÇÃO QUESTÃO 12 – RONDA 1	47
TABELA 4.5 – REFORMULAÇÃO QUESTÃO 15 – RONDA 1	47
TABELA 4.6 – REFORMULAÇÃO QUESTÃO 18 – RONDA 1	47
TABELA 4.7 – REFORMULAÇÃO QUESTÃO 23 – RONDA 1	48
TABELA 4.8 – REFORMULAÇÃO QUESTÃO 3 – RONDA 1	48
TABELA 4.9 – REFORMULAÇÃO QUESTÃO 8 – RONDA 1	48
TABELA 4.10 – REFORMULAÇÃO QUESTÃO 13 – RONDA 1	48
TABELA 4.11 – REFORMULAÇÃO QUESTÃO 16 – RONDA 1	49
TABELA 4.12 – REFORMULAÇÃO QUESTÃO 19 – RONDA 1	49
TABELA 4.13 – REFORMULAÇÃO QUESTÃO 20 – RONDA 1	49
TABELA 4.14 – REFORMULAÇÃO QUESTÃO 2 – RONDA 1	50
TABELA 4.15 – REFORMULAÇÃO QUESTÃO 5 – RONDA 1	50
TABELA 4.16 – REFORMULAÇÃO QUESTÃO 7 – RONDA 1	50
TABELA 4.17 – REFORMULAÇÃO QUESTÃO 21 – RONDA 1	50
TABELA 4.18 – REFORMULAÇÃO QUESTÃO 22 – RONDA 1	50
TABELA 4.19 – REFORMULAÇÃO QUESTÃO 6 – RONDA 1	51
TABELA 4.20 – REFORMULAÇÃO QUESTÃO 14 – RONDA 1	51
TABELA 4.21 – REFORMULAÇÃO QUESTÃO 17 – RONDA 1	51
TABELA 4.22 – REFORMULAÇÃO QUESTÃO 11 – RONDA 2.....	52
TABELA 4.23 – REFORMULAÇÃO QUESTÃO 17 – RONDA 2.....	52
TABELA 4.24 – REFORMULAÇÃO QUESTÃO 18 – RONDA 2.....	53
TABELA 4.25 – REFORMULAÇÃO QUESTÃO 19 – RONDA 2.....	53
TABELA 4.26 – REFORMULAÇÃO QUESTÃO 23 – RONDA 2.....	53
TABELA 4.27 – REFORMULAÇÃO QUESTÃO 17 – RONDA 3.....	54
TABELA 4.28 – DADOS SOCIO-DEMOGRÁFICOS AMOSTRA PRÉ-TESTE	55
TABELA 4.29 – RESULTADOS ARTIGO 1.....	56
TABELA 4.30 – RESULTADOS ARTIGO 2.....	57
TABELA 4.31 – RESULTADOS ARTIGO 3.....	58
TABELA 4.32 – RESULTADOS FORMULÁRIO DE COMPREENSÃO	59
TABELA 4.33 – RESULTADOS TEMPO DE PREENCHIMENTO	60
TABELA 5.1 – ESCALA DE PEDRO	64

Índice de figuras

ILUSTRAÇÃO 2.1 – ESTRUTURA ECR	12
ILUSTRAÇÃO 2.2 – <i>WEBPAGE PEDRO</i> – HOMEPAGE	19
ILUSTRAÇÃO 2.3 – <i>WEBPAGE PEDRO</i> – PESQUISA RÁPIDA	20
ILUSTRAÇÃO 2.4 – <i>WEBPAGE PEDRO</i> – PESQUISA AVANÇADA	20
ILUSTRAÇÃO 2.5 – <i>WEBPAGE PEDRO</i> – RESULTADOS PESQUISA RSS	21
ILUSTRAÇÃO 2.6 – <i>WEBPAGE PEDRO</i> – RESULTADOS PESQUISA ECRs	21
ILUSTRAÇÃO 2.7 – <i>WEBPAGE PEDRO</i> – INFORMAÇÃO DETALHADA DA PESQUISA	22
ILUSTRAÇÃO 3.1 – <i>WEBPAGE DELPHI DECISION AID</i> – FORMULAÇÃO QUESTÕES	37
ILUSTRAÇÃO 3.2 – <i>WEBPAGE DELPHI DECISION AID</i> – RESPOSTA DO PERITO	38
ILUSTRAÇÃO 3.3 – ESQUEMA DE TRADUÇÃO <i>PEDRO SCALE</i>	41
ILUSTRAÇÃO 4.1 – RESULTADOS RONDA 1 PROCESSO DELPHI	46
ILUSTRAÇÃO 4.2 – RESULTADOS RONDA 2 PROCESSO DELPHI	52
ILUSTRAÇÃO 4.3 – RESULTADOS RONDA 3 PROCESSO DELPHI	54
ILUSTRAÇÃO 4.4 – RESULTADOS RONDA 4 PROCESSO DELPHI	55
ILUSTRAÇÃO 4.5 – GRÁFICO DOS RESULTADOS FORMULÁRIO DE COMPREENSÃO	59
ILUSTRAÇÃO 6.1 – DESENHO DE ESTUDO REPRODUTIBILIDADE.....	68

Lista de Abreviaturas

CONSORT	Consolidated Standards of Reporting Trials
ECCs	Ensaio Clínicos Controlados
ECRs	Ensaio Clínicos Randomizados
DP	Desvio-padrão
PEDro	Physiotherapy Evidence Database
PBE	Prática Baseada na Evidência
PSVP_1	PEDro Scale Versão Portuguesa versão 1
PSVP_2	PEDro Scale Versão Portuguesa versão 2
PSRT	PEDro Scale Retrotradução
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
QUORUM	Quality of Reporting of Meta-analyses
RSs	Revisões Sistemáticas
WHO	World Health Organization

Capítulo I

Introdução

A prática baseada na evidência tem-se tornado uma das maiores preocupações de hoje em dia, quer para os profissionais de saúde quer para os utentes que usufruem da prestação de cuidados (12; 28). Não significa que se põe de parte a capacidade empírica do profissional, pelo contrário, essa capacidade será complementada com o que será objectivado por estudos de evidência, e até mesmo desencadeará o desenvolvimento de novas questões científicas (51).

Entende-se então por prática baseada na evidência o uso consciente, explícito e judicioso da melhor evidência existente resultante de investigação e estudos, disponíveis para influenciar na tomada de decisões sobre o tratamento a prestar aos doentes (12; 24). É uma abordagem para a tomada de decisão, tendo em conta a melhor evidência disponível, em parceria com o doente, decidindo no que melhor se adequa para o caso (28).

A Fisioterapia, como qualquer outra profissão da saúde, deve ter uma base científica assente na prática baseada na evidência. Tal facto envolve o uso de evidência proveniente de Ensaaios Clínicos Randomizados (ECRs), Ensaaios Clínicos Controlados (ECCs) e Revisões Sistemáticas (RSs), considerados os *golden standards* da evidência. No entanto, tais desenhos de estudo não são fáceis de concretizar na íntegra, por várias limitações relacionadas, por exemplo, com o acesso a amostras representativas e aleatorizadas, com *follow-ups* ou ainda com questões éticas. Por isso é necessário avaliar o rigor e consequente qualidade destes estudos, tornando-se imprescindível ter um instrumento fiável e que permita, de uma forma objectiva, criteriosa e pragmática analisar o rigor metodológico e científico dos estudos realizados (44; 46; 53).

A *PEDro scale* foi criada por Moseley et al em 1999 (46) e pretende, através de 11 itens, avaliar se foram seguidos os critérios de um ensaio clínico em condições ideais e consequentemente avaliar a qualidade do estudo na área da Fisioterapia (46; 51). Esta escala foi elaborada a partir da *Delphi list*, desenvolvida após um consenso de peritos em 1998 (72), que visava o mesmo objectivo, no entanto foi considerada insuficiente, por não contemplar parâmetros estatísticos. A necessidade da criação e validação de um instrumento deste tipo adveio da necessidade de catalogar os ensaios clínicos na base de dados *PEDro* (*Physiotherapy Evidence Database*) (48).

Desde então o referido instrumento de medida tem sido utilizado não só para a permitir aceder de uma forma mais prática à evidência existente no âmbito da Fisioterapia, como também se tornou num importante instrumento de uniformização na compilação de revisões sistemáticas, tanto na área específica da Fisioterapia (58) como nas restantes áreas profissionais (16).

A *PEDro Scale* torna-se então um bom instrumento na procura da melhor investigação para consolidar a decisão de tratamento. Este instrumento existe, até então, apenas em inglês (48).

Apesar da língua inglesa ser a língua universal na comunidade científica, tal facto constitui uma barreira para a compreensão, interpretação e aplicação do instrumento. É de elevada importância que o instrumento esteja traduzido e validado para as demais populações que o utilizam, neste caso, para a população portuguesa.

Para a tradução e adaptação de um instrumento com esta dimensão é necessário seguir uma série de etapas, já deliberadas em estudos anteriores (6; 41; 73), de forma a garantir a equivalência linguística do mesmo, para posterior averiguação das suas características psicométricas (validação). As traduções imediatas, como referem certos autores (29; 73), podem levar a erros na interpretação do instrumento, bem como do seu objecto de avaliação (41).

Por se tratar de uma área específica da fisioterapia, área de investigação, e conter termos técnicos muito específicos, a apreciação da tradução e contextualização cultural devem ser feitas por um painel de peritos (24). Para se proceder ao consenso da informação obtida, uma das melhores formas descritas na literatura é através do processo Delphi (14; 15; 27; 77).

O recurso à técnica Delphi permite estruturar o processo de comunicação de um grupo de indivíduos, como um todo, de forma a permitir que partilhem o debate de um assunto complexo, tendo como fonte de informação a opinião de peritos. O seu objectivo é a obtenção de uma opinião colectiva qualificada sobre determinadas questões, a partir de um grupo de pessoas seleccionadas (27; 77).

Este estudo pretende então traduzir e adaptar culturalmente a *PEDro scale* a fim de providenciar uma importante ferramenta na área da prática baseada na evidência na Fisioterapia, em Portugal. Constituirá assim uma medida importante para os fisioterapeutas portugueses que realizam investigação e para os que fundamentam a sua prática analisando de forma crítica a investigação publicada. Deixarão de ter uma barreira linguística.

Inicialmente pretendia-se averiguar também as características psicométricas, porém os autores da escala sugeriram que o estudo fosse decomposto em duas partes, uma vez para a validação seria necessário que a amostra fosse treinada na aplicação do instrumento.

Esta dissertação, realizada no âmbito da Tese de Mestrado em Ciências da Fisioterapia, obedece à estrutura a seguir descrita: no 1º capítulo procurou-se através da revisão da literatura fundamentar e contextualizar a pertinência do estudo, bem como a metodologia a ser utilizada; de seguida no capítulo da Metodologia são descritos os procedimentos bem como o desenho de estudo antes de se apresentar e discutir os Resultados; por fim procurou-se tirar conclusões apontando igualmente para as limitações deste estudo que justificam recomendações para estudos posteriores.

Capítulo II

Revisão da Literatura

Este capítulo procura fundamentar as bases programáticas que motivaram a questão orientadora central deste estudo bem como contextualizar e explicar as diferentes fases de desenho de estudos desta natureza.

2.1 Investigação e Prática Baseada na Evidência

O conhecimento adquire-se de muitas maneiras. Na maior parte das disciplinas e profissões, este foi adquirido através de diversas fontes no decurso da história: a intuição, as tradições e autoridade (crenças baseadas em costumes e tendências passadas), experiência pessoal (a tentativa e erro), raciocínio lógico (indutivo e dedutivo) e ainda a investigação científica. De todos os métodos de aquisição de conhecimento, a investigação é o mais rigoroso e, por conseguinte, o mais fidedigno (18).

A investigação científica pode ser definida como um processo de busca do conhecimento para estabelecer factos novos, resolver problemas (novos/existentes), desenvolver ou verificar novas ideias e teorias, utilizando geralmente o método científico. É um método particular de aquisição de conhecimentos, uma forma ordenada e sistemática de encontrar respostas para questões que necessitam duma investigação. Pelas suas funções, a investigação permite descrever, explicar e prever factos, acontecimentos ou fenómenos bem como descobrir, interpretar e desenvolver métodos e sistemas. O principal objectivo é o avanço do conhecimento humano numa ampla variedade de fenómenos, que se tornam assuntos científicos, do nosso mundo e universo (18).

Existem dois métodos fundamentais: o método de investigação qualitativa, onde o objectivo é mais descrever e interpretar do que avaliar; e o método de investigação quantitativo, que assenta na observação de factos objectivos e quantificáveis, tendo como finalidade a validação do conhecimento (18). Por permitir generalizar resultados, prever e controlar acontecimentos, o segundo método descrito é o mais utilizado na investigação clínica (27). Apesar do método quantitativo ser considerado mais fiável, actualmente defende-se que estes dois tipos de produção de informação complementam-se, sendo igualmente fiáveis quando o rigor metodológico está presente (32).

A investigação numa dada disciplina visa a produção de uma base científica para guiar a prática, assegurar a credibilidade de uma profissão, bem como definir os parâmetros e conteúdo funcional da mesma (18). No sector da saúde tem como objectivo fundamentar teoricamente a prática clínica, denominando-se a este processo de prática baseada na evidência (PBE) (24).

A prática baseada na evidência é uma abordagem onde a tomada de decisão tem em conta três tipos de informação: a investigação clínica, as preferências do utente e a experiência do profissional, levando à melhor opção possível (24; 30). Define-se como o uso consciente, explícito e judicioso da melhor evidência existente na tomada de decisões sobre os cuidados e/ou tratamentos a prestar ao utente, de forma individual. É a integração da capacidade empírica individual do profissional de saúde com a informação proveniente da investigação esquematizada (40). As indicações mais recentes referem que a PBE não deve provir da melhor evidência disponível, mas sim unicamente de investigação clínica de alta qualidade. Não significa que a restante investigação não tenha valor, mas que não deverá servir de base para o termo PBE (24).

A Fisioterapia, como qualquer outra profissão de saúde, apresenta uma base científica na PBE. Como descrito anteriormente, tal facto requer o uso de evidência proveniente essencialmente de RSs e ECs, a referida investigação clínica de alta qualidade (51). Apesar da prática da Fisioterapia ter-se baseado em parte, ao longo dos anos, na observação clínica e experiência profissional (24) tem-se verificado, ao longo desta última década, um crescimento exponencial de ECRs e ECCs na área (50).

A PBE em Fisioterapia deve envolver os seguintes passos (24):

1. Converter a necessidade da informação em questões orientadoras concretas (efeitos de intervenções, prognóstico, diagnóstico);
2. Pesquisar a melhor evidência com a qual se responderá às questões;
3. Analisar do ponto de vista crítico a evidência quanto à sua validade, impacto e aplicabilidade;
4. Integrar a evidência com a perícia clínica e com a especificidade dos pacientes;
5. Avaliar a eficácia e a eficiência da execução das etapas anteriores, procurando melhorar sempre (24).

A partir daqui, então, integrar o conhecimento na prática diária.

Por isso é necessário separar a investigação e qualificá-la da melhor forma, para que não se cometam erros por falta de rigor científico (62; 74), bem como classificar os diferentes tipos de estudos e a sua relevância científica (46; 72).

2.1.1 Níveis de evidência e graus de recomendação

Como referido, a investigação quantitativa é a mais utilizada na investigação clínica, e a Fisioterapia não é excepção.

A necessidade de uniformizar a informação na comunidade científica, e a falta de tempo dos profissionais para a filtragem de informação, levou a que, através da hierarquização dos tipos de estudo consoante a sua metodologia (tabela 2.2), se delimitasse níveis de evidência consoante a qualidade e tipo de estudo (tabela 2.2 e 2.3). Desta forma torna-se mais fácil e rápido aceder à melhor informação (28; 39).

TABELA 2.1 – HIERARQUIA DE EVIDÊNCIA CLÍNICA (39)

Método	
1.	Estudo analítico (teste de hipóteses)
	a) Ensaio clínico randomizado ou controlado (experimental);
	b) Estudo Coorte Prospectivo (observacional);
	c) Estudo Coorte Retrospectivo (observational);
	d) Estudo Caso-control (observational).
2.	Estudo Descritivo (formulação de hipóteses)
	a) Análise de tendências temporais (grupo, correlacional);
	b) Transversais (individual, num determinado momento);
	c) Series de casos (Grupo de pacientes não randomizado);
	d) Estudo de Caso (um só caso de uma patologia ou tratamento raro).
3.	Investigações utilizando sujeitos não-humanos ou modelos não vivos
	a) Estudos em modelos animais;
	b) Estudos em cadáveres, modelos sintetizados.

Existem várias classificações que delimitam os níveis de evidência (23; 39; 67; 69; 75; 76) e graus de evidência descritas na literatura. Em Fisioterapia, uma das utilizadas, é a referida na tabela 2.2, que tem em conta não só a metodologia utilizada como também a qualidade do estudo em questão.

TABELA 2.2 – NÍVEIS DE EVIDÊNCIA I (23)

Nível	Descrição
1++	Meta-análises de alta qualidade e/ou revisões sistemáticas de ECC e ECR *
1+	Baixo risco de viés em meta-análises e/ou revisões sistemáticas de ECC e ECR
1-	Elevado risco de viés em meta-análises e/ou revisões sistemáticas de ECC e ECR
2++	Revisões sistemáticas de alta qualidade de estudos de caso-control e/ou de cohort
2+	Baixo risco de viés em estudos de caso-control e/ou de cohort com relação causal
2-	Elevado risco de viés em estudos de caso-control e/ou de cohort sem relação causal
3	Estudos não-analíticos, i.e. relatórios de caso, estudos de caso, estudos de casos
4	Opinião empírica do perito

Esta diversificação nas hierarquias para a evidência provém também do facto de terem de se adequar ao âmbito do estudo (diagnóstico, prognóstico, intervenção, etc.) (28).

Outra hierarquia bastante citada é a enumerada na tabela 2.3, que está associada a graus de recomendação de PBE (tabela 2.4) (28).

No entanto, todas têm em comum o facto de colocarem as RSs em primeiro lugar, por constituírem uma redução do risco de viés. Apesar de também apresentarem limitações (24).

TABELA 2.3 – NÍVEIS DE EVIDÊNCIA II (28)

Nível	Descrição
Ia	Evidência a partir de Meta-análises e/ou Revisões sistemáticas de ECR *
Ib	Evidência a partir de pelo menos um ECR
IIa	Evidência a partir de pelo menos um estudo controlado sem randomização
IIb	Evidência a partir de pelo menos um estudo do tipo quasi-experimental
III	Evidência a partir de estudos descritivos não experimentais tais como estudos comparativos, correlacionais ou <i>case-control</i>
IV	Evidência a partir de relatórios de comité de peritos, opiniões ou experiência clínica de autoridades reconhecidas na área.

São os níveis de evidência científica que guiam os graus de recomendação de um determinado tratamento ou procedimento: quanto maior o nível, mais forte a ideia de que pode ser adoptado com segurança e benefício do paciente. Os graus de recomendação de evidência (tabela 2.4) são bastante utilizados nas RSs, pois estabelecem a relação entre a tipologia de estudo e a qualidade proveniente com o que deve ser em condições ideais o rigor metodológico (23).

TABELA 2.4 – GRAU DE RECOMENDAÇÃO DE EVIDÊNCIA (28)

Grau	Descrição
A	Directamente baseado no nível de evidência I
B	Directamente baseado no nível de evidência II ou em recomendações extrapoladas a partir do nível de evidência I
C	Directamente baseado no nível de evidência III ou em recomendações extrapoladas a partir do nível de evidência I ou II
D	Directamente baseado no nível de evidência II ou em recomendações extrapoladas a partir do nível de evidência I, II ou III

Esta abordagem classificativa apresenta limitações, devido a precisamente existirem vários critérios para definição de níveis de evidência (24). De qualquer forma é amplamente utilizada, nomeadamente no desenvolvimento de *guidelines* (28).

2.1.2 Ensaios clínicos

Os ensaios clínicos controlados randomizados constituem um dos principais avanços científicos entre os métodos de pesquisa durante o século XX, são considerados o padrão de referência nos métodos de investigação clínica (18).

2.1.2.1 Definição

Estudo clínico ou ensaio clínico controlado/randomizado é considerado a melhor fonte de evidência científica disponível e a melhor fonte de determinação da eficácia de uma intervenção. Trata-se de um modelo de pesquisa que compara o efeito e valor de uma intervenção (profilática ou terapêutica) de um grupo experimental (ou grupo teste) com um grupo de controlo em seres humanos. O primeiro é submetido à intervenção que se quer testar e o segundo recebe o tratamento convencional ou placebo. O investigador distribui o factor de intervenção a ser analisado de forma aleatória através da técnica de randomização; dessa forma, os grupos são formados por um processo aleatório de decisão (27; 56). Os grupos são acompanhados simultaneamente e prospectivamente. No final do período de tempo pré-estabelecido, os resultados dos dois grupos são analisados. Pode-se referir a fármacos, técnicas ou procedimentos. Permite assim verificar a eficácia da intervenção, medindo e comparando os diferentes desfechos depois da mesma (17).

O termo eficácia refere-se ao resultado de uma intervenção realizada sob condições ideais, bem controladas, como nos ensaios clínicos controlados. Efectividade refere-se ao resultado de uma intervenção aplicada sob as condições habituais da prática, que incluem as imperfeições de implementação que caracterizam o mundo quotidiano (18; 24).

Foi considerado que o termo “ensaio clínico”, de uma forma genérica, pode ser aplicado a qualquer forma de experiência planeada que envolva pacientes e que seja desenhada para elucidar o tratamento mais apropriado para futuros pacientes com uma dada condição médica. No entanto assim incluiria “ensaios clínicos não controlados” como um estudo em que todos os participantes recebem o factor de intervenção (65). Este seria apenas um estudo descritivo dos efeitos de uma intervenção num grupo, não sendo considerado tal estudo como um ensaio clínico, denominando no máximo de estudo quasi-experimental (18). Autores mais puristas guardam o termo ensaio clínico apenas para os ensaios controlados e randomizados, não aceitando nem o uso do termo para ensaios controlados, porém sem randomização (24; 28).

2.1.2.1 Metodologia

Em termos metodológicos, ao desenvolver um ensaio clínico o investigador começa por definir a população que se pretende estudar, que será constituída por casos de uma determinada doença, e escolhe uma amostra aleatória dessa população (18). Decompõe-se sempre a questão em 4 partes, utilizando a mnemónica PICO (24):

- **Patient or problem** (paciente ou problema);
- **Intervention or management strategy** (estratégia de intervenção ou gestão);
- **Comparative intervention** (intervenção comparativa);
- **Outcome** (resultado).

Na escolha da amostra é importante a definição de critérios de inclusão e/ou exclusão que tornem a amostra mais homogênea, aumentando a validade interna

do estudo, à custa, no entanto, de uma diminuição da capacidade de generalização (57).

Aos indivíduos dessa amostra é, então, de um modo aleatório, aplicada a intervenção que se pretende estudar ou uma intervenção alternativa, com a qual a primeira será comparada. São aleatoriamente definidos dois grupos: um que recebe a intervenção (o tratamento de que se pretende avaliar a eficácia) – *grupo experimental*; e um outro que recebe uma intervenção alternativa (um placebo, o tratamento convencional ou um outro tratamento alternativo) – *grupo controlo* (56). A atribuição aleatória da intervenção permite, em princípio, garantir que os dois grupos são homogêneos no início, excepto, no que toca à intervenção a ser avaliada. É importante, para evitar erros sistemáticos, que ao paciente seja, na medida do possível, ocultado o grupo a que pertence, uma vez que este conhecimento poderá influenciar o seu comportamento em relação à intervenção que lhe foi atribuída, para garantir que o participante seja cego ao estudo (55; 56).

Segue-se um período durante o qual as intervenções são aplicadas, os indivíduos são seguidos e a sua evolução clínica avaliada. No caso de a intervenção ser especificamente dada por uma classe profissional, como é o caso da Fisioterapia, também deve ser ocultado para os investigadores que aplicam a intervenção, para assim garantirem que quem administra a intervenção seja cego ao estudo (55).

São, depois, verificados os resultados esperados no grupo experimental e no de controlo. Também para evitar erros sistemáticos, ao investigador que avalia os resultados esperados também deve ser ocultado o grupo, a que o paciente pertence. Quando num ensaio clínico é ocultado o grupo a que o indivíduo pertence ao próprio doente e ao investigador que avalia o resultado final, o ensaio diz-se ser duplamente cego (17).

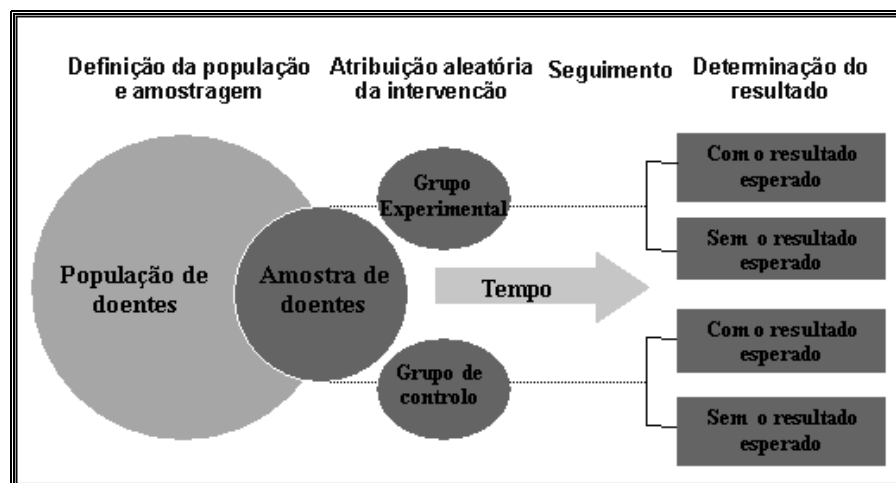


ILUSTRAÇÃO 2.1 – ESTRUTURA ECR (18; 54)

Finalmente, analisam-se as diferenças existentes entre os dois grupos definidos, no que diz respeito ao resultado final, tendo em conta, a randomização inicialmente feita (análise do tipo “intenção de tratamento”). Isto permitirá o cálculo de medidas de associação e impacto que permitem definir o efeito da intervenção em estudo na situação em causa (27). Devido ao processo utilizado, qualquer diferença existente entre os grupos é atribuída à intervenção que está a ser avaliada. Num ensaio clínico deve ser considerada, a possibilidade de existirem outros efeitos para além do efeito específico devido à intervenção em estudo (18). Estes outros efeitos designam-se por efeitos não específicos. Os efeitos não específicos podem ser de dois tipos: o efeito de Hawthorne, devido às alterações do comportamento dos participantes relacionadas, unicamente, com o facto de estarem a ser seguidos; e o efeito placebo, devido, unicamente, à aplicação de uma intervenção semelhante mas que não tem um mecanismo de acção específico conhecido. Além destes deve-se sempre considerar a possível existência de variáveis não controladas (18; 28).

Existem ainda formatos específicos de ECRs, como os denominados de *crossover* em que os sujeitos são distribuídos aleatoriamente para cada segmento do estudo em que cada segmento consiste numa sequência de dois ou mais tratamentos dados consecutivamente (42). O modelo mais simples é o estudo AB/BA, ou seja, os sujeitos colocados no segmento AB recebem primeiro o tratamento A e depois o B e no segmento BA ocorre a situação inversa. O objectivo dos estudos *crossover* é contrastar a resposta do mesmo sujeito a tratamentos diferentes. Tira assim o efeito da variação do sujeito, tornando assim este desenho de estudo potencialmente mais eficaz ao equiparado em que o sujeito é só exposto a um só tratamento. Em teoria os efeitos podem ser estimados com maior precisão tendo em conta o mesmo número de sujeitos. (42; 64)

2.1.2.3 Vantagens e Desvantagens

A grande vantagem dos estudos experimentais, nomeadamente os ensaios clínicos, é permitirem evitar, ou pelo menos controlar, erros sistemáticos através do processo de atribuição aleatória da intervenção, como foi já focado, e ocultando o objectivo de estudo a vários níveis (doente, fisioterapeuta, investigador e responsável pela avaliação do resultado esperado), promove-se a “cegueira” ao estudo. (65)

O ensaio clínico randomizado e duplamente cego é considerado o método padrão de avaliação das intervenções terapêuticas. É este o método que confere a melhor evidência de eficácia de um determinado tratamento e é, quase obrigatório ter este tipo de estudos como base para justificar a utilização de qualquer tratamento(16; 65).

Para além destas vantagens estes estudos, são estudos prospectivos, longitudinais, analíticos, pelo que têm como unidade de análise o indivíduo e como base de classificação dos participantes a existência de uma determinada exposição, neste caso, a intervenção em estudo. (18).

As principais desvantagens destes estudos são:

- Duração muito longa, existindo um grande período de tempo entre o início do estudo e o conhecimento dos resultados do estudo;
- Custo elevado devido ao longo período de seguimento e à apertada avaliação que é feita a todos os indivíduos;
- Pouca eficácia no estudo de doenças raras, uma vez que, nestes casos, se tornam mais longos e caros;
- Questões éticas que impedem, muitas vezes, que estudos experimentais possam ser feitos em humanos (18).

Para além das já apontadas, existem outras desvantagens ligadas a aspectos operacionais do próprio estudo e que poderão ser responsáveis pela existência de erros sistemáticos, nomeadamente, a existência de indivíduos que não aderem às intervenções que lhe foram atribuídas, a existência de "drop-out" (indivíduos que abandonam o estudo). É, também, um importante problema a impossibilidade de os sujeitos, fisioterapeutas e avaliadores serem cegos ao estudo(27).

Por último, deve referir-se que os ensaios clínicos são estudos que apresentam medidas da eficácia do tratamento no conjunto dos doentes analisados e, assim, eles não são uma garantia segura de eficácia em todos os doentes individualmente(74). Nos ensaios clínicos existe, por vezes, uma tão extensa selecção dos indivíduos a entrar no estudo, de modo a assegurar uma maior validade interna do mesmo, que a generalização a outros doentes pode ficar comprometida (27). Apesar de os ensaios clínicos serem a melhor razão para acreditar na eficácia de um tratamento, não há nada que ultrapasse a experiência do próprio doente com o tratamento (74).

2.1.2.4 Regras para exposição de Relatório - CONSORT

Um relatório de um ECR deveria comunicar da uma maneira transparente ao leitor, a razão pela qual o estudo foi realizado, e como foi conduzido e analisado. A uniformização da exposição dos relatórios tem sido uma preocupação crescente ao longo dos anos na comunidade científica (7; 43; 45; 61).

Em 1996, um grupo de investigadores (7), através de um consenso internacional, desenvolveu o CONSORT (Consolidated Standards of Reporting Trials), que inclui uma *checklist* e um fluxograma, para melhorar os relatórios sobre ECRs.

O padrão foi desenvolvido em resposta às evidências de que os ECRs têm sido relatados de forma inadequada durante as últimas três décadas, apesar dos esforços educacionais (68).

A lista identificava 21 elementos essenciais nos métodos, resultados e discussão necessária para avaliar a validade interna e externa de um relatório de ECR. O diagrama de fluxo contém informações sobre o progresso dos sujeitos através das diferentes etapas, incluindo o fluxo de participantes, de abandono (devido a que razões), e o calendário das medidas de desfecho primário e secundário (7). O CONSORT foi revisto e publicado em 2001 (45), continha uma lista de 22 itens e um fluxograma modificado, e mais recentemente em 2010 (61), com alteração só na *checklist*, agora com 25 itens.

Estas revisões surgiram porque, apesar desta última década ter melhorado muito em termos de qualidade na exposição científica (63), continuavam a existir estudos que estavam desadequadamente reportados (61).

Actualmente quase todos os jornais científicos exigem que o artigo de investigação cumpra os requisitos do CONSORT (9; 61; 68). A *checklist* e o fluxograma encontram-se disponíveis na página de internet da Organização CONSORT (2).

2.1.2.2 Contribuição para a PBE

O ECR é o modelo mais pragmático do método experimental em investigação clínica. O seu contributo será o maior uma vez que avalia o que determinada intervenção provocará directamente no objecto de estudo que deseja, ou seja, em seres humanos. (24; 65). Desta forma torna-se essencial este tipo de desenho para a fundamentação teórica de diagnóstico, prognóstico ou resultado de uma intervenção para a prática em Fisioterapia (40; 58). Não se tratando de uma “receita”, havendo necessidade de adequar às necessidades do utente, os resultados obtidos permitem ao fisioterapeuta desempenhar as suas funções com maior conforto e entendimento (24; 28; 40).

2.1.3 Revisões Sistemáticas

2.1.3.1 Definição

Uma revisão sistemática (RS) consiste numa compilação da informação de ECRs e ECCs relevantes, no sentido de minimizar vieses. Localiza, avalia e sintetiza toda a literatura disponível sobre um tópico, utilizando um desenho científico específico que tem que ser descrito na própria RS, factor que distingue este método, a RS, das outras revisões (1).

É recorrente confundir RS com os formatos a seguir discriminados:

- Revisões de literatura tradicionais (revisões narrativas de todo o tipo de informação sobre determinado tópico);
- Revisões integrativas (combinação de estudos quantitativos e qualitativos);
- Meta-análises (reanálise estatística de um conjunto de dados referente a ECRs com o mesmo objecto de estudo);
- Revisões qualitativas (revisões de estudos qualitativos) (19; 20; 60).

A RS é diferente, trata-se da análise e síntese sistemática de estudos com evidência (classe I a classe III). Além de limitar o viés, melhora a fidedignidade e precisão das conclusões (59).

É particularmente útil em convergir um número de estudos, que por vezes obtêm resultados contrários, e sintetizar os seus resultados. Assim são também utilizadas para definir se os achados científicos são consistentes e generalizáveis pelas populações, contextos e variações de tratamento, ou se os achados variam significativamente em subgrupos particulares (1; 19).

Por este motivo RSs de ECRs são consideradas o mais alto nível de evidência para avaliar a eficácia das intervenções (1).

2.1.3.2 Metodologias Actuais

A *Cochrane Collaboration* é uma cooperação internacional que tem como objectivo ajudar profissionais e utentes a tomarem decisões bem informadas sobre cuidados de saúde, bem como na preparação, manutenção e promoção da acessibilidade de RSs sobre os efeitos de intervenções na saúde. Hoje em dia é a cooperação que mais RSs promove e supervisiona. É também a que descreve a estrutura e etapas de uma RS (66).

Metodologicamente uma RS deve apresentar a seguinte estrutura:

- 1) Introdução
- 2) Objectivos
- 3) Métodos da revisão
- 4) Resultados
- 5) Conclusão e discussão.

E deve seguir as seguintes etapas:

- 1) Formulação do problema ;
- 2) Localização e selecção dos estudos;
- 3) Apreciação crítica dos estudos;
- 4) Recolha de dados;
- 5) Análise e apresentação de resultados;
- 6) Interpretação dos resultados;
- 7) Sugestão de melhoria e actualização das revisões. (70)

Existem outras metodologias, mas são muito próximas desta descrição (59). As RSs na *Physiotherapy Evidence Database* seguem esta metodologia (47).

Para a avaliação dos artigos são utilizadas escalas de avaliação de ECRs, entre as quais a *PEDro Scale*. (13)

2.1.3.2 Regras para exposição de relatório de RS - QUORUM - PRISMA

O *QUORUM statement* (Quality of Reporting of Meta-analyses) era uma *checklist* de parâmetros com o objectivo de melhorar a qualidade dos relatórios das RSs, à semelhança do que o CONSORT faz para os ECRs (9; 43). Como em qualquer investigação, a qualidade das RSs irá depender do rigor metodológico com que o autor conduziu o estudo. Para que o processo seja mais transparente e uniforme o QUORUM estabelecia uma *checklist* de 21 itens para assegurar a inclusão da informação necessária para avaliar a qualidade da RS (9; 10). O fluxograma compreendia a informação detalhada do fluxo dos ECRs ao longo do processo da RS.

O termo meta-análise encontrava-se aqui introduzido porque, confusamente, na América do Norte as RSs são denominadas de meta-análises, enquanto na Europa o termo é utilizado para a combinação estatística dos resultados dos estudos que compõem uma RS (24; 52).

Por esse motivo, em 2006, reuniu-se um grupo de peritos para definir melhor a terminologia e rever o *QUORUM*. Nessa altura decidiu-se então modificar o nome para *PRISMA* (Preferred Reporting Items for Systematic reviews and Meta-Analyses) (33). Esta ficou composta por 26 itens com um novo fluxograma. Esta nova versão, como o nome indica, avalia RSs e Meta-análises, mas como conceitos distintos (33).

A *checklist* e o fluxograma encontram-se disponíveis no website da organização (34).

2.1.3.3 Contribuição para a PBE

A compreensão do que são RSs e como implementá-las na prática torna-se, actualmente, obrigatória para todos os profissionais envolvidos na prestação de cuidados de saúde. (66)

A maioria das revisões sistemáticas é sobre eficácia de uma determinada técnica ou terapia. No entanto, as RSs de testes de diagnóstico é também de importância superior, e afecta directamente a eficácia dos cuidados de saúde (24).

Os resultados das RSs podem ser utilizados em muitas populações e contextos diferentes ao redor do mundo. Os revisores devem tentar ajudar o leitor a aplicar os resultados da sua revisão a casos individuais (59).

As revisões sistemáticas são necessárias para a integração eficiente de informações válidas e fornecer uma base para tomada de decisão racional. O uso de métodos sistemáticos e explícitos nas revisões limita a possibilidade de viés (erros sistemáticos) e reduz os efeitos do acaso, proporcionando resultados mais confiáveis, a partir dos quais se pode tirar conclusões e tomar decisões (60). Permite também a criação de *guidelines*, linhas orientadoras para a conduta de determinado campo de intervenção

Assim, uma revisão sistemática é o nível mais alto de evidência tornando-se num instrumento muito útil e importante. Ajuda o fisioterapeuta e o paciente a tomar a melhor decisão (24).

2.1.4 Instrumentos de Medida de Ensaios Clínicos

Como já foi verificado, existe uma necessidade em avaliar o rigor metodológico dos ensaios clínicos. Quer na abordagem por níveis de evidência, quer na compilação nas RSs de ECs (13; 44), é imprescindível a existência de uma ferramenta que permita avaliar e qualificar os ECs (37; 53).

2.1.4.1 Descrição e relevância

Como todos os outros instrumentos de medida, este tipo de instrumento propõe avaliar um determinado objecto de estudo: o rigor da metodologia científica inerente aos ECRs/ECCs. Para então extrapolar os resultados mais significativos para a prática clínica, ou até mesmo compilar um grupo dos mesmos estudos numa RS (37; 71). Numa revisão feita por Olivo em 2008 (53), verificou-se que existem muitos instrumentos que têm esse objecto de estudo, porém nem todos têm todos apresentam todas as características psicométricas avaliadas.

2.1.4.2 Aplicações

Como já verificado é importante que se constate o rigor da metodologia científica inerente aos estudos científicos, para que se possa confiar nos resultados daí provenientes. Aplica-se também nas RSs na selecção dos ensaios clínicos, para que esta seja feita de uma forma sistemática e prática, e assim permitir uma melhor filtragem da informação (60).

2.1.4.3 Exemplos

Como exemplos estão descritas na tabela 2.5 algumas escalas de medida de qualidade de ECRs. Algumas são específicas para a área da Fisioterapia, outras transversais aos ensaios clínicos e outros ainda especificamente de outras áreas de saúde (53).

TABELA 2.5 – ESCALAS DE AVALIAÇÃO DE ECRs (53)

	Descrição	Área de Aplicação
Fisioterapia	Jadad Scale	ECRs Dor
	Maastricht Scale	ECRs Fisioterapia
	Delphi list	ECRs
	PEDro Scale	ECRs Fisioterapia
	Maastricht Amsterdam List	ECRs Raquialgias
	Maastricht Amsterdam List Modified	ECRs Raquialgias
	Van Tulder Scale	ECRs Raquialgias
	Bizzini Scale	ECRs Dor patelo-femoral
Outras Áreas da Saúde	Chalmers Scale	ECRs Farmacológicos
	Reisch Scale	ECRs Terapêuticos
	Tyson Scale	ECRs Terapêuticos
	Andrew Scale	ECRs Meios de contraste
	Andrew Modified Scale	ECRs Meios de contraste radiográficos
	Balas Scale	ECRs Serviços de Saúde
	Downs and Black Scale	ECRs Saúde Pública
	Arrivé Scale	ECRs Exames radiológicos
	Yates Scale	ECRs Psicológicos para a dor

O CONSORT, já descrito acima, é considerado por esta autora (53) também como um instrumento de avaliação mas os autores descrevem-no mais como uma lista de regras para formulação de relatório (61).

Na área da fisioterapia a escala mais indicada, em termos gerais, é a *PEDro Scale* (53).

2.2 Physiotherapy Evidence Database

Em 1999, o Centro de Fisioterapia Baseada em Evidências no *The George Institute For Global Health*, em Sidney, Austrália, desenvolveu uma base de dados electrónica (ilustração 2.2) de investigação clínica de alta qualidade, ECRs e RSs (28). Esta base de dados providencia detalhes bibliográficos e resumos de ECRs, RSs e linhas orientadoras (*guidelines*) em Fisioterapia. (25; 26; 28)



ILUSTRAÇÃO 2.2 – WEBPAGE PEDRO – HOMEPAGE (48)

Esta base de dados é composta, na verdade, por quatro páginas:

- PEDro pesquisa rápida (Ilustração 2.3) e pesquisa avançada (Ilustração 2.4) - para profissionais de saúde,
- Allied Health Evidence - em protocolo com outros grupos profissionais;
- Physiotherapy choices – para os utentes de Fisioterapia (48).

Na pesquisa avançada (Ilustração 2.4) pode-se fazer a busca por método de estudo, abordagem terapêutica, problema clínico, parte do corpo ou por subespecialidade ou ainda por pontuação (no caso dos ECRs) (28), além dos já normais parâmetros de resumo, autor, título, fonte e data (24).

Apresenta uma limitação significativa, pois não é possível executar buscas com E e OU, tendo que se decidir por um dos vocábulos (24). Outra limitação prende-se ao facto de não dividir as citações no âmbito de diagnóstico, prognóstico e resultados em Fisioterapia (28).

The screenshot shows the PEDro website's 'Simple Search' page. At the top, the PEDro logo and 'PHYSIOTHERAPY EVIDENCE DATABASE' are visible. Navigation links include Home, New Search (Simple), New Search (Advanced), and Search Help. A font size adjustment tool is in the top right. The main section is titled 'Simple Search' and contains a single text input field for 'Search term (or terms):' and a 'Search' button. The footer features the logo of 'THE GEORGE INSTITUTE for Global Health' and its affiliation with 'THE UNIVERSITY OF SYDNEY'.

ILUSTRAÇÃO 2.3 – WEBPAGE PEDRO – PESQUISA RÁPIDA (48)

The screenshot shows the PEDro website's 'Advanced Search' page. It includes the same header and navigation as the simple search page. The 'Advanced Search' section contains multiple search criteria: 'Abstract & Title:', 'Therapy:', 'Problem:', 'Body Part:', 'Subdiscipline:', 'Method:', 'Author/Association:', 'Title Only:', and 'Source:'. It also includes date filters for 'Published Since:' and 'New records added since:', a 'Score of at least:' field, and a 'Return:' dropdown set to '20 records at a time'. Under 'When Searching:', there are radio buttons for 'Match all search terms (AND)' (selected) and 'Match any search term (OR)'. A 'Start Search' button is at the bottom. The footer is identical to the simple search page.

ILUSTRAÇÃO 2.4 – WEBPAGE PEDRO – PESQUISA AVANÇADA (48)

Como vantagem apresenta a oportunidade de submeter críticas, caso não se concorde com as pontuações, e é atualizada mensalmente (28; 48).

Para ser elegível para a PEDro, o relatório tem que satisfazer 5 critérios:

- O estudo tem que comparar pelo menos 2 intervenções (ou uma intervenção com controlo ou placebo);
- Pelo menos uma das intervenções tem que ser utilizada na prática da Fisioterapia;
- As intervenções avaliadas têm que ser aplicadas a participantes que são representativos da população a que se destina a técnica de intervenção;
- A distribuição dos participantes tem que ser aleatória;
- O relatório tem que estar publicado num jornal científico *peer-reviewed* (26).

Relativamente aos resultados da pesquisa são apresentados como está descrito nas ilustrações 2.5 e 2.6. As revisões sistemáticas aparecem primeiro e não têm pontuação, pois para esse caso não é aplicável (ilustração 2.5) (26).

Title	Method	Score (/10)	Select Record
Support surfaces for pressure ulcer prevention (Cochrane review) [with consumer summary]	systematic review	N/A	Select
The influence of the therapist-patient relationship on treatment outcome in physical rehabilitation: a systematic review [with consumer summary]	systematic review	N/A	Select
Cupping for stroke rehabilitation: a systematic review	systematic review	N/A	Select
The effect of balance training on balance performance in individuals poststroke: a systematic review	systematic review	N/A	Select
A systematic review of common physiotherapy interventions in school-aged children with cerebral palsy	systematic review	N/A	Select
Reported quality of randomized controlled trials in neglect rehabilitation	systematic review	N/A	Select
Real-time kinematic, temporospatial, and kinetic biofeedback during gait retraining in patients: a systematic review [with consumer summary]	systematic review	N/A	Select

ILUSTRAÇÃO 2.5 – WEBPAGE PEDRO – RESULTADOS PESQUISA RSS (48)

A seguir os ECRs que apresentam logo na listagem a pontuação obtida na *PEDro Scale* (assinalado na ilustração 2.6), estando ordenados da pontuação mais alta para a mais baixa. Por isso a informação organiza-se do maior nível de evidência para o mais baixo (24; 28; 36).

Title	Method	Score (/10)	Select Record
Motor learning principles for rehabilitation: a pilot randomized controlled study in poststroke patients	clinical trial	8/10	Select
Long-term effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized controlled trial comparing two different physical therapy interventions	clinical trial	8/10	Select
Does cueing training improve physical activity in patients with Parkinson's disease?	clinical trial	7/10	Select
Open, randomized pilot study after first stroke: a 3.5-year follow-up	clinical trial	6/10	Select
Impaired leg vasodilatory function after stroke: adaptations with treadmill exercise training	clinical trial	4/10	Select
Effects of therapeutic modalities on patients with post stroke shoulder pain	clinical trial	4/10	Select

ILUSTRAÇÃO 2.6 – WEBPAGE PEDRO – RESULTADOS PESQUISA ECRs (48)


A informação detalhada de cada estudo (ECR ou RS) compreende os dados gerais do artigo (nome dos autores, título, fonte) e também o método, a

página de internet onde se encontra e ainda a pontuação na *PEDro scale* pormenorizada (ilustração 2.7) (24).

Relativamente à pontuação tem sido apontadas muitas críticas, principalmente porque não há um intervalo de pontuação descrito como melhor evidência (28), no entanto não deixa de ser um processo útil para filtrar a informação decorrente dos ECRs (51). Deve-se só ter em atenção e avaliar os ECRs antes de tomar a decisão/opção de tratamento e não confiar unicamente no resultado da pontuação. É preciso bom senso na adequação ao paciente(28).

Como já foi focado para catalogar os ECRs nesta base de dados foi necessário o desenvolvimento de uma escala que permitisse, de uma forma objectiva, avaliar os ECRs. Dai a obtenção da *PEDro Scale* (38; 46).

Detailed Search Results	
Author/Association:	Piron L, Turolla A, Agostini M, Zucconi CS, Ventura L, Tonin P, Dam M
Title:	Motor learning principles for rehabilitation: a pilot randomized controlled study in poststroke patients
Source:	Neurorehabilitation and Neural Repair 2010 Jul-Aug;24(6):501-508
Method:	clinical trial
Method Score:	8/10 [Eligibility criteria: Yes; Random allocation: Yes; Concealed allocation: Yes; Baseline comparability: Yes; Blind subjects: No; Blind therapists: No; Blind assessors: Yes; Adequate follow-up: Yes; Intention-to-treat analysis: Yes; Between-group comparisons: Yes; Point estimates and variability: Yes. Note: Eligibility criteria item does not contribute to total score] *This score has been confirmed*
Abstract:	Copyright release for this abstract has not been obtained. Abstract and full text may be available at: http://nnr.sagepub.com/content/24/6/501

 THE GEORGE INSTITUTE
for Global Health


 Affiliated with
THE UNIVERSITY OF
SYDNEY

ILUSTRAÇÃO 2.7 – WEBPAGE PEDRO – INFORMAÇÃO DETALHADA DA PESQUISA (48)

2.3 PEDro Scale

2.3.1 Descrição do Instrumento

A *PEDro Scale* baseia-se na lista de *Delphi*, desenvolvida por Verhagen et al (72). A lista, na sua maior parte, baseia-se num “consenso de peritos” e não em dados empíricos. Incluíram-se na *PEDro Scale* dois itens adicionais, que não constavam da lista de *Delphi* (os itens 8 e 10 da *PEDro Scale*). À medida que forem disponibilizados mais dados empíricos, tornou-se possível ponderar os itens da escala para que a pontuação obtida a partir da aplicação da *PEDro Scale* reflectisse a importância de cada um dos itens da escala (46; 51).

2.3.2 Fundamentos e objectivos

O objectivo da *PEDro Scale* consiste em auxiliar os utilizadores da base de dados PEDro a identificar rapidamente quais dos ensaios clínicos randomizados (ECR ou ECC) arquivados na base de dados PEDro poderão ter validade interna (critérios 2-9), e poderão conter suficiente informação estatística para que os seus resultados possam ser interpretados (critérios 10-11). Um critério adicional (critério 1) que relaciona a validade externa (ou “potencial de generalização” ou “aplicabilidade” do ensaio) foi conservado para que a lista de Delphi esteja completa, mas este critério não será usado para calcular a pontuação *PEDro* apresentada no endereço *PEDro* na internet (46; 51).

Esta escala também é utilizada para filtrar resultados e avaliar ECR e ECC, e para compilar revisões sistemáticas. (53)

Esta provado também que é mais sensível que outras escalas com a mesmo objectivo, nomeadamente a *Jadad Scale* (8).

2.3.3 Escalas, conceitos e pontuações

A *PEDro scale* (anexo I) é então uma *checklist* com 11 itens (tabela 2.6) para ECR ECC. Esta escala consiste na verificação de 11 critérios. (26)

Os ECR E ECC que preencham os critérios de inclusão da *PEDro* são avaliados quanto á sua qualidade utilizando a *PEDro scale*. A classificação é dada pelo número de critérios cumpridos (excepto o primeiro). As orientações para a utilização do instrumento especificam que serão então os critérios 2-11 os utilizados para fins de pontuação, a fim de se obter uma pontuação de 10 (26).

TABELA 2.6 – PEDRO SCALE ORIGINAL (26)

Critéria	Description	Yes/No
1.	Specification of eligibility criteria	
2.	Random allocation.	
3.	Concealed allocation.	
4.	Prognostic similarity at baseline.	
5.	Subjects blinding.	
6.	Therapist blinding.	
7.	Assessor blinding.	
8.	Greater than 85% follow-up for at least one key outcome.	
9.	Intention-to-treat analysis.	
10.	Between group statistical analyses for at least one key outcome.	
11.	Points estimates of variability provided for at least one key outcome.	

2.3.4 Administração e aceitabilidade

É um instrumento prático, porém é necessário conhecimento subjacente à metodologia científica. É utilizado principalmente por fisioterapeutas (46), mas também por outros profissionais (16).

Para efeitos de RS ou outros estudos científicos este instrumento tem que ser utilizado por um examinador treinado (com os requisitos exigidos pelo *PEDro steering group*) (38).

Não está descrito o tempo de preenchimento (53).

2.3.5 Fiabilidade e validade

A *PEDro scale* tem um alto nível de validade de conteúdo, sendo desenvolvida a partir da *Delphi list* e demonstrou alta fiabilidade inter-observador (Kappa=0,88) (46).

Quanto a validade ela foi recentemente avaliada em dois estudos (11; 35). No primeiro estudo, os investigadores testaram a validade de constructo com um modelo Rasch e concluíram que é uma medida válida ($r=0,99$) (11). Também foi considerado válida a soma dos itens para obter a pontuação da *PEDro scale*, assim como esta pontuação tem que ser tratada como uma escala de intervalo-nível (11). No segundo estudo, investigou-se a validade de constructo e de convergência comparando a escala com a *Jadad Scale* e a de *Van Tulder*. A escala apresentou validade de constructo e convergência para o valor total e para 8 dos 10 itens. (35)

Apesar de apresentar algumas desvantagens, como o facto de não contemplar a dimensão da amostra, a adequação dos testes estatísticos e a presença ou não de normalidade, este instrumento consegue avaliar bem o rigor do estudo (28).

2.3.6 Campos de aplicação

É utilizada na *webpage* e na compilação de revisões sistemáticas. É utilizada na prática clínica também. (13; 25)

A escala PEDro não deverá ser usada como uma medida da “validade” das conclusões de um estudo, pois apenas avalia o rigor metodológico do estudo. A escala não deve ser utilizada para comparar a “qualidade” de ensaios realizados em diferentes áreas de terapia, porque nalgumas áreas da prática da fisioterapia não é possível satisfazer todos os itens da escala (49).

2.4 Tradução e adaptação cultural de um instrumento

2.4.1 Descrição e relevância

Como a maioria dos instrumentos de medida (questionários, escalas, testes, etc...) a *PEDro Scale* apresentava-se, até então, apenas na língua original, a inglesa (48). Existem vários factores a ter em conta na tradução de um instrumento para que mantenha as características psicométricas iniciais da sua língua original para a língua alvo (termo utilizado para a língua para a qual o instrumento será traduzido) (6).

Durante a tradução de um instrumento devem ser avaliados alguns aspectos relativos à equivalência linguística:

- Equivalência semântica – avaliação da equivalência gramatical e de vocabulário;
- Equivalência idiomática – verificar se expressões idiomáticas e coloquialismos são substituídos por expressões equivalentes;
- Equivalência conceptual – avaliar a equivalência de conteúdo, verificando se transmite o mesmo conceito;
- Equivalência cultural – verificar se os termos utilizados são coerentes com as experiências vividas pela população à qual se destina, dentro do seu contexto cultural. (27; 73)

Para a tradução e adaptação cultural devem-se então seguir uma serie de etapas que vêm descritas na literatura (6).

2.4.2 Etapas descritas

Após revisão sistemática da literatura, os autores Guillemín et al (1993) (22), analisando trabalhos que abordaram alguma metodologia de tradução e adaptação cultural, propuseram um conjunto de instruções padronizadas para a adaptação transcultural de instrumentos de medida. Essas instruções, mais conhecidas por etapas, foram revistas e actualizadas pelos mesmos autores em 2000(6) e por Wild et al (2005) (73) mais recentemente.

2.4.1.1 Preparação

Nesta fase o autor deve ter delineado toda a metodologia. Deve preparar o grupo de pessoas com que vai trabalhar:

- Seleccionar os tradutores;
- Contactar e pedir de autorização formal aos autores do instrumento;
- Convidar os autores do instrumento a participarem no estudo;
- Recrutar ou reunir informação de pessoas chave na população alvo, da área de estudo do instrumento a traduzir (73).

2.4.1.2 Tradução

A tradução deve ser feita, em condições ideais, por dois tradutores bilingues cuja língua materna é a língua alvo (Português), que produzem duas versões traduzidas (73). No entanto está descrito que pode existir apenas um tradutor uma vez que a 1ª versão será posteriormente revista por peritos (18).

2.4.1.3 Reconciliação - Processo Delphi

A reconciliação pode ser feita de três formas: revisão de um painel de investigadores, revisão por uma pessoa alheia ao estudo e ainda pelos tradutores que produziram a primeira versão (73).

Neste estudo a reconciliação será feita por um painel de peritos que analisa as ou a versão adicionando críticas e sugestões produzindo a 2ª versão. Este processo pode ser feito através da técnica Delphi que será descrita pormenorizadamente mais à frente (*ver processo Delphi*).

2.4.1.4 Retrotradução

A retrotradução é a tradução da 2ª versão na língua alvo para a língua original que deve ser feita por dois tradutores bilingues cuja língua materna é a língua original do instrumento, em condições ideais. Tal como a tradução, uma vez que vai ser revista *a posteriori*, o estudo descreve que pode ser apenas um tradutor (73).

2.4.1.5 Relatório retrotradução vs versão original

Nesta fase é feita uma comparação da versão retrotraduzida e da versão original para salientar e investigar possíveis discrepâncias entre o original e a versão traduzida. O estudo de Wild et al (73) refere que este factor não é consensual, mas está descrito que este relatório pode ser realizado por alguém não específico do país origem, mas sempre que possível deve ser feito pelos autores do instrumento original, uma vez que mais do que ninguém, conhecem o seu valor semântico. A versão retrotraduzida tem que ser linguisticamente equivalente à versão original. Se não for o processo anterior terá que ser repetido (73).

2.4.1.6 Harmonização

Esta etapa compreende a harmonização de várias versões (línguas e populações). É um passo nem sempre utilizado, por se apresentar pouco claro quanto ao seu objectivo. É referido que uma das hipóteses seria promover uma reunião entre várias pessoas chave das diversas versões para obter a harmonização. Por esta etapa estar mal definida, necessitando de mais investigação, esta não fará parte do desenho de estudo (73).

2.4.1.7 Pré-teste

Uma vez que a versão retrotraduzida da 2ª versão é considerada equivalente, esta é aplicada num pequeno grupo, que compõe o pré-teste, no sentido de testar palavras alternativas e para verificar o nível de compreensão, interpretação e relevância cultural da tradução.

Um pré-teste é a aplicação de um questionário, na sua versão preliminar, a uma amostra de indivíduos, com o objectivo de identificar perguntas-problema que justifiquem uma modificação da redacção, alteração do formato ou mesmo serem eliminadas da versão final (27).

Os participantes primeiro preenchem o instrumento e depois, através de entrevista ou de formulário de compreensão, avaliam qualitativamente a escala, verificando-se se entenderam o significado das questões formuladas, se encontraram dificuldades na interpretação das mesmas e se responderam adequadamente. O grupo deve ser pequeno, de 10 a 30 sujeitos (73).

2.4.1.8 Análise resultados pré-teste

Realiza-se então a análise dos resultados dos participantes comparando a interpretação da 2ª versão para salientar e corrigir discrepâncias, compondo a 3ª e última versão da língua alvo. As questões com mais de 15 % de respondentes com dúvidas devem ser revistas pelo painel de peritos e reaplicadas noutra grupo de participantes (73).

2.4.1.9 Revisão final

Revisão final da 3ª versão para corrigir algum erro tipográfico, gramatical ou outros (73).

2.4.1.10 Versão final e Relatório

O relatório da versão final do instrumento é então redigido no fim, documentando o desenvolvimento de cada fase da tradução. Estando concluída esta etapa pode-se proceder ao processo de validação (73).

2.5 Validação de um instrumento

2.5.1 Descrição e relevância

Após o pré-teste, o investigador deve aplicar o instrumento numa amostra maior e verificar então as propriedades psicométricas da versão final do instrumento (41).

Existem vários autores que se reportam ao número total que deverá conter a amostra que aplica ou será aplicado o instrumento (6; 21; 41). No entanto a maioria dos autores sugere que sejam pelo menos dez participantes por item do instrumento (6).

Ao adaptar um instrumento ou uma escala e submetê-la a traduções e retro-traduções, é preciso assegurar-se que a escala ainda mantém as propriedades psicométricas testadas na versão original. Para isso é necessário revalidar o instrumento analisando a fiabilidade e validade do mesmo, através de testes específicos, estando alguns descritos na tabela 2.7 (6).

TABELA 2.7 – TIPOS DE FIABILIDADE E VALIDADE (6)

	Tipo	Descrição	Testes estatísticos
Fiabilidade	Consistência Interna ou Homogeneidade	Correlação entre os diferentes itens do questionário e entre os itens e o somatório total	Chronbach's alpha
	Fiabilidade Intra-examinador	É o grau de concordância entre avaliadores da avaliação de um teste ou questionário.	Intraclass Correlation Coefficient
	Fiabilidade Inter-examinador	É o grau de concordância entre as múltiplas repetições de um teste ou questionário, realizadas pelo mesmo avaliador.	Kappa de Cohen Kappa de Fleiss
Validade	Validade de face	Validade aparente ou validade externa, indica se, na sua aparência, o instrumento parece estar avaliando as qualidades desejadas.	São garantidas pela adaptação cultural, obtida normalmente pelo painel de peritos e pelo pré-teste -
	Validade de conteúdo	Relata se o conceito tem conteúdo ou domínios importantes	
	Validade de constructo (convergente ou divergente)	Se a medida correlaciona bem com outras medidas do mesmo fenómeno	Coeficiente de Correlação de Pearson
	Validade de critério (preditiva ou concorrente)	É a habilidade do instrumento medir o que reclama medir Comparação com padrão de ouro.	

À excepção da validação de face, os itens anteriores não foram explorados porque a validação deste instrumento não será feita neste estudo, por se necessitar de uma amostra que esteja apta a utilizar o mesmo. Para isso será necessário um treino específico que será efectuado pelo *PEDro steering group*.

2.6 Processo Delphi

2.6.1 Descrição

O recurso à técnica Delphi permite estruturar o processo de comunicação de um grupo de indivíduos, como um todo, de forma a permitir que partilhem o debate de um assunto complexo, tendo como fonte de informação a opinião de peritos. O seu objectivo é a obtenção de uma opinião colectiva qualificada sobre determinadas questões, a partir de um grupo de pessoas seleccionadas. Tal abordagem é indicada para situações onde não existem informações precisas ou dados históricos, ou em situações onde se deseja estimular novas ideias e consolidar antigas. As ideias que emergem têm em conta os conhecimentos, a experiência e a visão estratégica de julgar e prever mudanças a longo prazo dos peritos, no sentido de “prever o futuro” e/ou intervir na sua configuração. Partindo

das opiniões, pontos de acordo e divergências entre os peritos, pretende-se chegar a certos elementos de consenso e a uma hierarquização de aspectos importantes e secundários relativos ao problema em estudo. As principais características da técnica Delphi consistem na utilização de um painel de peritos para obter consenso, sustentado no facto de os participantes não terem confrontação frente a frente, na garantia de anonimato das respostas e no uso de ferramentas estatísticas simples para identificar níveis de concordância (27).

A esta técnica associam-se as seguintes características básicas (27; 77):

- Anonimato: a opinião dos membros do grupo é obtida através de um questionário, pelo que os participantes não interagem directamente uns com os outros;
- *Feedback*: os resultados são enviados aos participantes para que estes validem novamente as suas opiniões. O objectivo básico do *feedback* é distribuir todas as informações disponíveis e geradas pelo grupo entre os seus participantes;
- Interação: a técnica consiste em sucessivas rondas cuja quantidade é determinada de acordo com um critério de consenso do grupo ou estabilização dos julgamentos individuais;
- Resposta estatística do grupo: a opinião do grupo é definida, no final de cada ronda, através da agregação das opiniões individuais.

Na prática, este “*feedback*” controlado consiste em remeter a todos os participantes, em rondas sucessivas, a informação que se vai gerando em cada etapa. É neste sentido que o investigador, recorrendo a esta estratégia, vai transmitindo aos participantes a informação livre de interferências, eliminando aquela que considera não relevante, repetida e confusa, com o objectivo de enviar uma mensagem clara (tanto em conteúdo, como na forma e linguagem), que agrupe a opinião do grupo sobre a área em discussão. (77) Desta forma, comunicando aos participantes um resumo da etapa precedente, reduz o “ruído semântico”, fornecendo ao grupo somente aquilo que se refere aos objectivos e metas do estudo, evitando que os participantes se desviem dos pontos centrais do problema (14; 15).

O consenso é o resultado de um processo de interação em grupo em que os participantes têm a oportunidade de trocar informações entre si sobre o assunto em estudo até atingirem o nível de acordo previamente fixo ou estabelecido pelo investigador (14).

O método Delphi é baseado num processo estruturado para colectar e refinar o conhecimento de um grupo de especialistas por meio de uma série de questionários, intercaladas com comentários de opinião. A técnica Delphi representa um dispositivo de comunicação útil entre um grupo de especialistas e, portanto, facilita a formação de uma decisão de grupo. Sublinha-se a importância do método Delphi como técnica de exploração monovariável para a previsão de tecnologia. Afirma-se ainda que o método Delphi foi desenvolvido a fim de tornar possível a discussão entre especialistas, sem permitir um certo comportamento social interactivo, como acontece durante uma discussão em grupo normal e dificulta a formação da opinião. Na falta de conhecimentos científicos completos, tem que se confiar na própria intuição ou seja na opinião de especialistas. O

método Delphi tem sido amplamente utilizado para gerar previsões em tecnologia, educação e outros campos (14).

2.6.2 Objectivo

O principal objectivo do processo ou técnica de Delphi é chegar a um consenso de opinião profissional, relativamente a qualquer tópico. O método é particularmente útil quando têm que ser tomadas decisões em situações em que haja demasiada informação ou informação a menos, ou em que haja provas contraditórias. É um meio através do qual o conhecimento e a experiência profissional relevante podem ser sintetizados e utilizados para se chegar a uma decisão fundamental (77).

Não havendo informação publicada ou empírica adequada, a opinião de conceituados especialistas, podem ser obtidas com recurso ao processo Delphi, como um meio através do qual o conhecimento e experiência profissional relevante pode ser sintetizada e utilizada para se chegar a uma decisão fundamentada. Embora as reuniões de comités e os grupos de convergência também possam ter esta finalidade, os seus resultados dependem, em grande parte, das interações de grupos que têm lugar, acontecendo muitas vezes ser a pessoa mais poderosa, dominadora ou influente a ditar a discussão e conclusões (77). Por seu turno, a técnica de Delphi recorre a uma abordagem assente num levantamento por correspondência individual (correio ou electrónica (3)) que visa especialistas de vulto na matéria. Deste modo, evitam-se eventuais influências suscitadas pela interacção entre os membros do grupo, numa reunião, do que resulta muitas vezes um conjunto de respostas mais sincero e, por conseguinte, mais valioso (14).

2.6.3 Método

No essencial, a abordagem de Delphi envolve a identificação de profissionais especialistas na área a debater, sendo-lhes então enviado um formulário de levantamento, cujo propósito é apurar os seus pontos de vista acerca dos tópicos a ser investigados. Tal pode ser atingido de qualquer das seguintes formas (77):

- Convidando os especialistas a darem as suas opiniões sobre o tópico, num formato de resposta livre e não estruturada, ou
- O investigador traça uma perspectiva sobre o tópico ou sobre os temas centrais sob investigação, e pede ao painel de especialistas que expressem o seu grau de concordância com a mesma recorrendo a uma escala numérica (escala de *Likert*).

O formulário de levantamento completo é devolvido ao investigador, que examina as respostas, compila uma lista de pontos de vista mais defendidos pelos especialistas, redistribuindo depois essa lista reduzida ao mesmo painel (14).

Aos especialistas/peritos é pedido que ordenem os itens da lista, pela ordem de importância que atribuem a cada um, e as respostas são novamente devolvidas ao investigador que procederá a uma nova filtragem. O procedimento é

repetido numerosas vezes até que se obtenha uma pequena lista de itens que reúna a concordância de todos os especialistas que compõem o painel (77).

A concordância ou consenso, no processo Delphi, assume duas formas. Em primeiro lugar, a concordância dos especialistas relativamente às questões investigadas. Em segundo lugar, a concordância entre os membros do painel (77).

Ao longo de todo o processo as respostas são anónimas e confidenciais, evitando-se, deste modo, as influências inter-pessoais que ocorrem nos encontros frente a frente (77).

2.6.4 Procedimento

Na prática pode-se dividir esta técnica em 11 etapas descritas (27; 77) abaixo.

- Etapa 1: Identificar o painel de peritos;
- Etapa 2: Determinar a vontade dos peritos em servir o painel;
- Etapa 3: Reunir o contributo individual do tema específico e depois compila-lo numa instrução básica;
- Etapa 4: Analisar os dados provenientes do painel;
- Etapa 5: Compilar a informação para uma nova ronda;
- Etapa 6: Analisar a nova informação (repetir esta etapa até haver consenso) e devolver os dados finais ao painel;
- Etapa 7: Solicitar a cada membro do painel opinião sobre a recolha e síntese de dados.
- Etapa 8: Analisar a informação, sintetizar e partilhar com o painel;
- Etapa 9: Possibilidade de fazer o tratamento estatístico com o Coeficiente de Kendall para obter concordância ou então por métodos de estatística descritiva, descrevendo percentagens. Só pode ser efectuado quando os peritos avaliam quantitativamente as questões ou a pertinência das mesmas (27; 77).

2.6.5 Vantagens

O facto da técnica de Delphi ser um processo anónimo, encoraja respostas imunes a influências e a pressões de colegas e de cima para baixo. Permite também que indivíduos bem informados, mas menos influentes e inibidos, expressem as suas opiniões, o que muitas vezes é difícil numa reunião aberta. Como resultado, os dados obtidos podem ser considerados mais honestos que aqueles obtidos por métodos comparativos (14; 77).

Como a abordagem é feita através de levantamento por correio ou via electrónica, o que significa que é pouco dispendiosa. O método por correio também assegura que as respostas dos participantes não são “contaminadas” por pressão do entrevistador (27).

O método possibilita a recolha de grande quantidade de dados, de um vasto leque de especialistas que podem ter “backgrounds” distintos. Seria

impossível, por variadas razões de ordem prática e financeira, obter o mesmo nível e quantidade de dados, a partir de entrevistas ou reuniões (27).

O método, ao englobar um leque de opiniões de peritos, garante a validade do seu conteúdo (15).

É uma técnica altamente flexível, que pode ser adaptada a um grupo de disciplinas temáticas e profissionais, bem como a vários desenhos de estudo (15; 77).

2.6.6 Desvantagens

Esta técnica tem sido no entanto, alvo de críticas quanto à falta de rigor científico, principalmente no processo de amostragem; inexistência de linhas de recomendação quanto aos critérios de inclusão para a selecção do painel de especialistas e por estar dependente da colaboração activa dos mesmos (77).

Neste contexto o estudo tem como objectivo traduzir a *PEDro Scale*, obtendo a equivalência linguística e adaptando culturalmente à população portuguesa.

Capítulo III

Metodologia

Este capítulo define a estratégia metodológica utilizada e os materiais necessários.

Este estudo estaria subdividido em duas partes, primeiro na tradução e adaptação cultural e posteriormente na validação. No entanto, como já referido, a segunda parte será feita em conjunto com o *PEDro Steering Group*. Este estudo refere-se ao processo de tradução, que seguiu uma sequência específica.

3.1 Desenho de Estudo

Este estudo consiste na tradução e adaptação cultural da *PEDro Scale* (Anexo I).

O estudo seguiu uma série de etapas, descritas abaixo, promovendo a equivalência linguística e a adaptação cultural do instrumento para a sua população alvo – essencialmente fisioterapeutas, mas também para a restante comunidade científica.

Este trabalho enquadra-se numa metodologia de estudo transversal descritivo.

3.2 Objectivo de estudo

Este estudo tem como objectivo traduzir e adaptar a medida de ensaios clínicos em fisioterapia *PEDro Scale*.

3.3 Procedimentos da Tradução e Adaptação Cultural

Conforme enumerado na revisão da literatura deste manuscrito, existem etapas pré-descritas de forma a garantir uma tradução correcta. O estudo descrito seguiu os procedimentos de Beaton et al (6) e Wild et al (73). Foi inserido o processo Delphi, ajustando o parâmetro de reconciliação, por se tratar de uma temática muito específica (investigação). Pretendeu-se assim reenforçar a equivalência conceptual (por definição de termos chave por parte dos peritos).

3.3.1 Preparação

Após a concepção do desenho de estudo, pediu-se autorização ao *PEDro steering group* (Anne Moseley, Chris Maher, Rob Herbert, Cathie Sherrington and Mark Elkins), autores da *PEDro scale*, para traduzir, adaptar e validar o referido instrumento para a cultura portuguesa (Apêndice I). A autorização foi obtida com aprovação por unanimidade (Anexo II) após discussão sobre o desenho de estudo e recepção de relatório inicial com a descrição das etapas de todo o processo (Apêndice II). Porém sugeriram que o estudo fosse decomposto em duas partes, uma vez que para o processo de validação será necessário que a amostra esteja treinada na aplicação do instrumento. Este treino é dado por eles através de um *training package*(48). O grupo acompanhou todo o processo, com esclarecimentos e sugestões, que também compõem este trabalho.

Nesta fase também foi feito o levantamento de todos os possíveis peritos na área de investigação em Fisioterapia.

Foram também recrutadas as tradutoras.

3.3.2 Tradução

Foi efectuada a primeira tradução de inglês para português por uma tradutora bilingue cuja língua materna é a portuguesa (tradutora 1). Esta versão foi averiguada pela investigadora e pela tradutora, designando-se por PEDro Scale versão portuguesa 1 (PSVP_1 – Apêndice III).

3.3.3 Reconciliação – Processo Delphi

Para a etapa da reconciliação relativa à tradução deve-se comparar e combinar mais de uma opinião de um perito sobre a mesma. Este processo foi realizado através da técnica Delphi, descrita na literatura como a melhor para obter consenso de um grupo de peritos.

3.3.3.1 Selecção do painel de peritos

Por se tratar da área de investigação científica, foi seleccionado um painel de especialistas, por conveniência, que cumpriram os seguintes requisitos:

- Fisioterapeuta;
- Grau académico pelo menos de Mestre;
- Docentes ensino superior em Fisioterapia;
- Língua materna portuguesa.

Foram seleccionados 10 peritos dos quais 6 acordaram participar.

3.3.3.2 Recolha de dados peritos

Como método foi utilizado a *webpage Delphi Decision Aid* (3), que permite executar o processo Delphi de uma forma informatizada, esquematizada e anónima. Esta página da Internet permitiu formular e corrigir as questões e respectiva escala (ilustração 3.1), enviar o formulário das questões por e-mail para todos os peritos, fazendo automaticamente a estatística inerente à recolha das

avaliações (numero total de respostas, média, desvio-padrão e mediana), à medida que os peritos foram respondendo às questões e/ou fazendo correcções e sugestões. No fim de cada ronda era feito um sumário das respostas.

Delphi
forecastingprinciples.com

Log out

Authorization:
• Administrator log in
• Expert log in

Related useful links and texts:
• Forecasting website
• >>> What's New <<<
• Long-Range Forecasting: Judgemental Methods
• The Delphi Technique as a forecasting tool

Session name: PEDro Scale tradução
Round #: 1
Round state: Question Designing

View other rounds:
[Round 1]

Question Designing / Expert Selection / Process Monitoring / Results Review

New Question Start Round Back to Session List Log Out

Text	Type	Features	
Concorda com a tradução do item 1 da escala de PEDro "Os critérios de elegibilidade foram especificados?" do original "eligibility criteria were specified"?	Scaled	Scale: 0 (Discordo totalmente) to 5 (Concordo totalmente) No opinion allowed: no	edit delete

If satisfied with your questions, move on to selecting experts.

ILUSTRAÇÃO 3.1 – WEBPAGE DELPHI DECISION AID – FORMULAÇÃO DE QUESTÕES (3)

Foi uma ferramenta importante, que só pode ser utilizada em questões onde o resultado seja em rankings ou estimativas quantitativas (3).

3.3.3.3 Critérios de avaliação

Neste estudo as perguntas foram colocadas de forma a conseguir uma aprovação numa escala de *Likert* de 0 a 5, sendo que 0 é discordo totalmente e 5 concordo totalmente.

Exemplo:

Concorda com a tradução do item 1 da *PEDro Scale* “xxxx” do original “xxxx”?

A seguir à questão o perito tinha uma escala de *Likert* para preencher e um campo vazio para formular correcções e/ou sugestões (Ilustração 3.2).

O definido como critério de consenso foi o valor 5 atribuído por 5 em 6 dos peritos em todas as questões colocadas.

Pretendeu-se que os peritos colocassem sugestões e correcções, no número de rondas necessárias até que houvesse consenso na tradução da escala.

A seguir a cada ronda, a investigadora procedeu às modificações consoante as correcções e/ou sugestões, procedendo a uma nova ronda até que as questões fossem consensuais entre os peritos.

Question:

Concorda com a tradução do item 1 da escala de PEDro "Os critérios de elegibilidade foram especificados?" do original "eligibility criteria were specified." ?

Please answer this question on the following scale:
0 (Discordo totalmente) to 5 (Concordo totalmente)

Please fill in your answer here: *

Please provide any additional comments/justification for your answer in the field below:

Submit answer

ILUSTRAÇÃO 3.2 – WEBPAGE DELPHI DECISION AID – RESPOSTA DO PERITO (3)

Foram necessárias 4 rondas (Apêndice IV) para o consenso (*ver cap. resultados*) de pelo menos 5 dos 6 peritos em todos os itens. Obteve-se então a PEDro Scale versão portuguesa 2 (PSVP_2 – Apêndice V).

3.3.4 Retrotradução

Foi feita uma retro tradução da PSVP_2 para língua de origem (apêndice VI), de português para inglês, por outra tradutora bilingue cuja língua materna é o inglês, a *PEDro Scale* retro tradução (PSRT). Foi feita uma retrotradução mais literal, por se achar mais adequada à natureza deste instrumento.

3.3.5 Retrotradução vs original

A autora Anne Moseley e o seu aluno Leonardo Costa compararam a PSRT com a versão original a fim de verificar se o instrumento era semanticamente idêntico. Encontraram alguns termos diferentes, específicos da comunidade científica, que foram considerados semanticamente idênticos (Tabela 3.1).

TABELA 3.1 – EQUIVALÊNCIA DOS TERMOS DIFERENTES DA RETROTRADUÇÃO

Original	Retrotradução	Equivalência
"...at baseline..."	"Initially..."	<input checked="" type="checkbox"/>
"...blinding of..."	"...at blind fashion"	<input checked="" type="checkbox"/>
"...intention-to-treat..."	"...treatment intention..."	<input checked="" type="checkbox"/>
"...between-group comparisons..."	"...intergroup statistical comparison..."	<input checked="" type="checkbox"/>
"...Point mesures..."	"...Accuracy measurement..."	<input checked="" type="checkbox"/>

Outros termos e frases mais elaboradas, como também reforçam os autores, prendem-se ao facto da língua portuguesa ser mais rica em palavras e sinónimos que a língua inglesa.

A retrotradução foi considerada, na sua generalidade, semanticamente idêntica ao original (Anexo II).

Com este passo concluído, procedeu-se à aplicação na população alvo da PSVP_2 a fim de verificar o nível de compreensão e a existência de frases e termos problema.

3.3.6 Pré-teste

Procedeu-se ao pré-teste. Como indicação para este passo os autores da *PEDro Scale* definiram os seguintes factores:

- Utilização de três ECRs registados e já avaliados na *PEDro database* (dois dos sessenta existentes em Português e um em inglês);
- Amostra com sujeitos não treinados especificamente.

Pretendeu-se assim, além de aferir a clareza e compreensão do instrumento na população alvo, avaliar a capacidade de utilização do instrumento por sujeitos não treinados, verificando assim a relevância e importância do treino.

Pretendeu-se também perceber se havia diferenças na avaliação de artigos na língua inglesa relativamente à portuguesa.

3.3.6.1 Seleccção dos participantes no estudo

A amostra foi recolhida por conveniência, tendo como população alvo fisioterapeutas.

Os critérios de inclusão para a amostra do pré-teste foram:

- Fisioterapeutas Licenciados;
- Consentimento informado para colaborar no estudo (Apêndice VII).

3.6.6.2 Protocolo

Os participantes avaliaram três artigos de ECRs com o auxílio do PSVP_F. O primeiro (31) e o terceiro (5) são escritos em português (Brasil) e o segundo (4) em inglês. A escolha dos artigos foi aleatória: retirou-se 2 entre os 60 portugueses e 1 entre os 9000 em língua inglesa.

Primeiramente preencheram a folha de rosto com os respectivos dados sócio-demográficos. Posteriormente responderam ao formulário de compreensão (Apêndice VIII) indicando numa escala de *likert* (0-Incompreensível e 5-Claro) o nível de compreensão de cada item, pontuação final e instruções do referido instrumento. Para cada parâmetro referido tiveram um espaço para resposta aberta.

3.3.6.3 Recolha de dados

A recolha dos dados foi feita em sala, uma vez que é um processo demorado e foi necessário verificar a média do tempo de resposta.

3.3.7 Análise resultados Pré-teste

A análise dos resultados foi feita de duas formas:

- Análise estatística - utilizou-se medidas de localização e de tendência central para caracterização dos participantes no estudo, e na comparação com a avaliação do *website* (nível de concordância em percentagem);
- Análise de conteúdo – verificou-se a frequência das sugestões e dúvidas para analisar se diziam respeito a falta de conhecimento do instrumento ou relativamente à clareza do mesmo.

Foi definido como valor de aprovação para a compreensão e clareza dos parâmetros o valor ≥ 4 (compreensível a claro), por média aritmética, com um desvio padrão de no máximo ± 1 (algumas dúvidas a claro), uma vez que, como referido na literatura, o instrumento tem que ser compreensível.

Quanto à concordância do valor dado pelo pré-teste em comparação com o já descrito no *website* não foi definido nenhum valor específico, uma vez que serviu para levantamento de dados.

No entanto seria considerado um valor aceitável o nível de concordância $\geq 80\%$, com um desvio-padrão entre respostas entre $\pm 0,5$.

Os resultados permitiram compor a terceira e última versão da Escala de PEDro, que foi praticamente idêntica à PCVP_2.

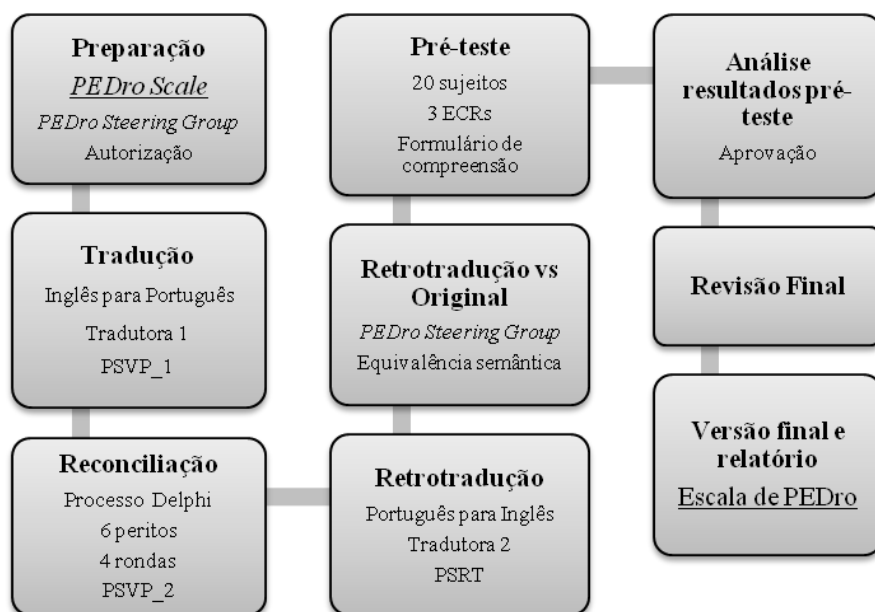
3.3.8 Revisão final

Após os resultados do pré-teste fez-se uma revisão final da 3ª versão para corrigir erros ortográficos, tipográficos, gramaticais ou outros. Não se encontrou porém nenhum erro.

3.3.9 Versão Final e Relatório

Obteve-se a PEDro Scale versão portuguesa final (PSVP_F) – Escala de PEDro (Apêndice IX). O relatório da versão final do instrumento foi redigido, documentando o desenvolvimento de cada fase da tradução (Apêndice X), compondo esta dissertação.

A título de resumo demonstra-se o processo que a tradução seguiu na sequência representada na ilustração 3.3.

ILUSTRAÇÃO 3.3 – ESQUEMA DE TRADUÇÃO *PEDRO SCALE*

3.3.10 Validação

A validade de face foi garantida pelo painel de peritos e pelo pré-teste, porém a restante validação não foi feita porque fará parte de um processo realizado em conjunto com o *PEDro Steering group* (ver Cap. VI).

Capítulo IV

Resultados

4.1 Tradução e retrotradução

A tradução e retrotradução foram executadas por duas tradutoras bilingues conforme o referido na metodologia. O resultado da comparação da versão original e a retrotradução feita pelo *PEDro steering group* revelou que eram equivalentes (Anexo III).

4.2 Adaptação Cultural

A adaptação cultural para equivalência semântica foi obtida através do consenso de peritos no processo Delphi e do formulário de compreensão no pré-teste.

4.2.1 Processo Delphi

No processo Delphi as rondas foram realizadas por exclusão das perguntas consensuais e reformulação das não consensuais com o auxílio das correções e sugestões dos peritos (Apêndice IV).

4.2.1.1. Caracterização do painel de peritos

O painel de peritos foi composto por 6 fisioterapeutas (4 ♀ e 2 ♂). Todos com grau académico de Mestre e todos docentes de ensino superior em Fisioterapia. A língua materna de todos é o português.

4.2.1.2 Ronda 1

Na ronda 1 apenas as questões 1 e 10 foram consensuais, necessitando todas as outras de reformulação (Ilustração 4.1). Conforme descreve o gráfico, o consenso foi definido com pelo menos 5 em 6 atribuições do valor máximo da escala de *Likert*.

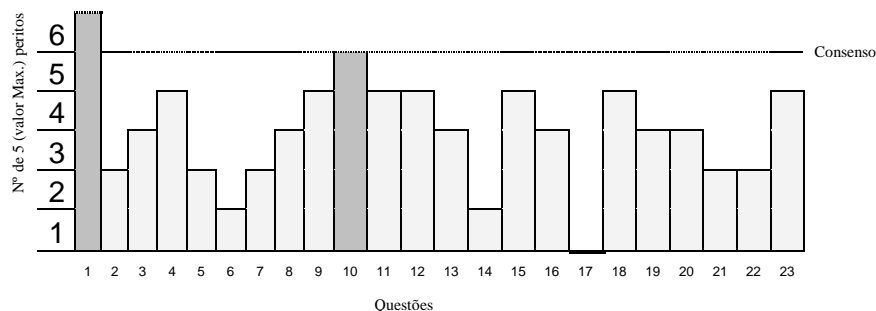
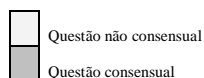


ILUSTRAÇÃO 4.1 – RESULTADOS RONDA 1 PROCESSO DELPHI

As questões 4, 9, 11, 12, 15, 18 e 23 obtiveram a aprovação máxima de 4 peritos (4/6), no entanto apresentavam alguns termos que não foram considerados correctos e/ou indicados pela maioria.

Na questão 4 foram reformuladas as frases apresentadas na tabela 4.1, para melhorar a fluência da frase.

TABELA 4.1 – REFORMULAÇÃO QUESTÃO 4 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...à partida...”	“Inicialmente, ...”	1
“...aos mais importantes indicadores de prognóstico.”	“...aos indicadores de prognóstico mais importantes.”	1

Na questão 9 o termo “intenção de tratamento” foi aceite, o que se pensava mais problemático. A questão foi reformulada para melhorar a semântica no primeiro caso, e a conceptualidade no segundo (tabela 4.2).

TABELA 4.2 – REFORMULAÇÃO QUESTÃO 9 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...para os quais...”	“...a partir dos quais ...”	2
“...como previsto...”	“...conforme a distribuição...”	1

Na questão 11, o termo “Point measure and variability”, não foram consideradas equivalentes a “pontos de medição e medidas de variação”. Foram sugeridos “ponto de medida e medidas de variabilidade” e “medidas únicas e medidas de variabilidade”. Por alguma dúvida em torno destes termos, após revisão de termos estatísticos, adoptou-se os termos referidos na tabela 4.3.

TABELA 4.3 – REFORMULAÇÃO QUESTÃO 11 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...pontos de medição e medidas de variação...”	“...tanto medidas de localização como medidas de variabilidade...”	0

Na questão 12 surgiram só correcções de construção frásica (tabela 4.4).

TABELA 4.4 – REFORMULAÇÃO QUESTÃO 12 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...tornar-se-á possível...”	“...pode vir a ser possível...”	1
“...relaciona a validade externa...”	“...que diz respeito à validade externa...”	1
“...de que o tratamento tenha sido clinicamente útil.”	“...de que o tratamento seja clinicamente útil...”	1
“...efeitos positivos do mesmo suplantam os negativos...”	“...se os efeitos positivos superam os negativos...”	1
“...e também aferir a relação de custo-eficácia do tratamento.”	“...e aferir a relação de custo-eficácia do tratamento.”	1

Na questão 15 (tabela 4.5) surgiram correcções relativas a construção frásica mas também de equivalência conceptual (na terceira).

TABELA 4.5 – REFORMULAÇÃO QUESTÃO 15 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...se o relatório refere...”	“...se o relatório referir...”	2
“...ser explicitado.”	“...ser explícito.”	3
“Sorteios tais ...”	“Procedimentos tais ...”	2
“...registo no hospital...”	“...registo hospitalar...”	2

Na questão 18 o termo “outcome measure” não foi considerado equivalente do termo “indicador de eficácia”, por isso prende-se a equivalência semântica e conceptual (tabela 4.6).

TABELA 4.6 – REFORMULAÇÃO QUESTÃO 18 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...ser usada como indicador de eficácia.”	“...ser usada como medida de resultados.”	3

Na questão 23 o termo “point measure” não foi considerado equivalente ao termo “ponto de medida” pelo que, assim como na questão 11, após revisão de termos estatístico utilizou-se o termo referido na tabela 4.7.

TABELA 4.7 – REFORMULAÇÃO QUESTÃO 23 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...um ponto de medida...”	“...uma medida de localização ...”	0
“...critério foi satisfeito....”	“...critério foi cumprido...”	1

As questões 3, 8, 13, 16, 19 e 20 obtiveram a aprovação máxima de 3 peritos (3/6) apresentando na altura alguns termos considerados incorrectos e algumas frases cuja construção também foi considerada incorrecta ou inadequada.

Na questão 3 a frase inteira foi reformulada por uma questão semântica e conceptualidade (tabela 4.8).

TABELA 4.8 – REFORMULAÇÃO QUESTÃO 3 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“A distribuição pelos grupos foi sigilosa.”	“A distribuição dos sujeitos foi cega.”	3

Na questão 8 a tradução da frase referida na tabela 4.9 foi considerada incorrecta.

TABELA 4.9 – REFORMULAÇÃO QUESTÃO 8 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...foram efectuadas para mais...”	“...foram obtidas em mais ...”	4

A questão 13 foi reformulada nos parâmetros indicados na tabela 4.10, devido a factores de construção frásica e equivalência semântica.

TABELA 4.10 – REFORMULAÇÃO QUESTÃO 13 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“Só se atribui pontuação...”	“A pontuação só será atribuída...”	3
“...inequivocamente satisfeito.”	“...claramente satisfeito.”	3
“...for possível considerar que o...”	“...existir a possibilidade de um...”	2

Na questão 16, foi reformulada pelos mesmos motivos da questão 13, nos dois primeiros casos e por reforço semântico no último (Tabela 4.11).

TABELA 4.11 – REFORMULAÇÃO QUESTÃO 16 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“Distribuição sigilosa...”	“Distribuição cega, ...”	2
“...a distribuição foi sigilosa...”	“...a distribuição foi cega...”	2
“...envelopes fechados...”	“...envelopes opacos fechados...”	1

Na questão 19 os termos “Blind” “Blinding” não foram consensuais em termos de semântica e conceito, não foram considerados equivalentes a “estudo cego em relação a...” pelo que foi sugerido a reformulação indicada na tabela 4.12.

TABELA 4.12 – REFORMULAÇÃO QUESTÃO 19 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“Estudo cego em relação ao sujeito, terapeuta ou avaliador...”	“Estar cego ao estudo...”	3
“O estudo só é cego em relação aos sujeitos e aos terapeutas...”	“Mais ainda, sujeitos e terapeutas só são considerados “cegos”...”	3
“...o estudo é considerado cego em relação ao avaliador se o for em relação ao sujeito.”	“...o avaliador é considerado “cego” se o sujeito foi “cego””	3

A questão 20 foi reformulada apenas para melhorar a adequação e fluência frásica no primeiro e reforço do conceito no segundo (Tabela 4.13).

TABELA 4.13 – REFORMULAÇÃO QUESTÃO 20 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...para os quais se obtiveram ...”	“...a partir dos quais se obtiveram...”	3
“...diferentes momentos...”	“...diferentes momentos no tempo...”	2

As questões 2, 5, 7, 21, 22 obtiveram a aprovação máxima apenas de 2 peritos (2/6) apresentando na altura alguns termos que não foram considerados correctos, algumas frases cuja construção também não estava correcta e alguns termos que necessitavam de reestruturação para melhor compreensão dos termos/conceitos e ideias.

Na questão 2 o termo “estudo transversal” não é equivalente a “estudo *crossover*” não existindo, no entanto, nenhuma tradução específica para este termo (talvez “estudo cruzado”, porém não é consensual na literatura portuguesa). O segundo parâmetro indicado na tabela 4.14 foi reformulado para melhor compreensão da frase.

TABELA 4.14 – REFORMULAÇÃO QUESTÃO 2 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...estudo transversal ...”	“...estudo <i>crossover</i> ...”	5
“...foram aleatoriamente ordenados para receber tratamento).”	“...foram colocados em grupos de forma aleatória de acordo com o tratamento recebido).”	3

As questões 5 e 7 foram reformuladas na íntegra (tabelas 4.15 e 4.16), pelos mesmos motivos da já descrita questão 19.

TABELA 4.15 – REFORMULAÇÃO QUESTÃO 5 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“Estudo cego em relação aos sujeitos.”	“Todos os sujeitos participaram de forma cega no estudo.”	4

TABELA 4.16 – REFORMULAÇÃO QUESTÃO 7 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“Estudo cego em relação aos avaliadores que mediram pelo menos um resultado-chave”	“Todos os avaliadores que mediram pelo menos um resultado-chave, fizeram-no de forma cega.”	5

Na questão 21, as frases referidas na tabela 4.17 foram reformuladas para melhorar a compreensão e intensificar a equivalência semântica.

TABELA 4.17 – REFORMULAÇÃO QUESTÃO 21 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...tal como previsto ...”	“...conforme o grupo atribuído...”	3
“...receberam o tratamento ou condição de controlo previstos.”	“...receberam o tratamento ou condição de controlo, conforme a distribuição por grupos.”	2

Na questão 22 o termo “inter-grupal” foi considerado desadequado e não muito utilizado na comunidade científica portuguesa pelo que foi sugerido o termo indicado na tabela 4.18. Relativamente à segunda alteração esta diz respeito a um termo que melhor se adequa à ideia que se pretende passar e não que estivesse incorrecto.

TABELA 4.18 – REFORMULAÇÃO QUESTÃO 22 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“Uma comparação inter-grupal estatística ...”	“Uma comparação estatística inter-grupos ...”	5
“...da mudança num grupo em relação à mudança no outro...”	“...das alterações num grupo em relação às alterações no outro ...”	2

As questões 6 e 14 obtiveram a aprovação máxima de 1 perito (1/6).

Na questão 6, a frase foi toda reformulada pela razão já descrita na questão 9 (tabela 4.19).

TABELA 4.19 – REFORMULAÇÃO QUESTÃO 6 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“Estudo cego em relação aos terapeutas que administraram a terapia.”	“Todos os terapeutas que administraram a terapia fizeram-no de forma cega”	5

A questão 14 foi reformulada para melhor adequação frásica (Tabela 4.20).

TABELA 4.20 – REFORMULAÇÃO QUESTÃO 14 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...as origens...”	“...a origem...”	2
“...lista de critérios...”	“...lista de requisitos...”	1

A questão 17 foi a única que não obteve a aprovação máxima de nenhum dos peritos (0/6), tendo críticas relativas à equivalência semântica e conceptual (Tabela 4.21).

TABELA 4.21 – REFORMULAÇÃO QUESTÃO 17 – RONDA 1

Ronda Anterior	Reformulação	Nº de peritos
“...medida do estado da condição ...”	“...medida de severidade da condição ...”	1
“...diferenças de resultados para os diversos grupos.”	“...diferenças clinicamente significativas de resultados para os diversos grupos.”	1

Procedeu-se ao envio para nova ronda após reformulação.

4.2.1.3 Ronda 2

Na ronda 2 as questões 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 16, 20, 21 e 22 foram consensuais (Ilustração 4.2).

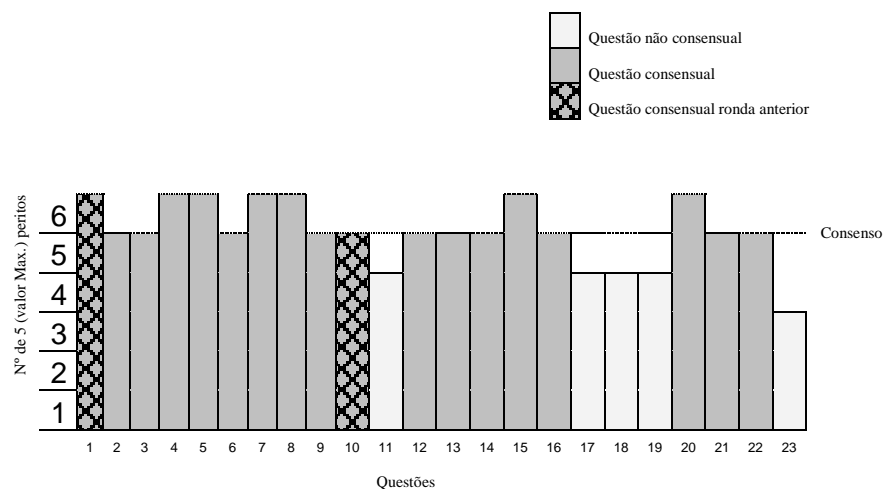


ILUSTRAÇÃO 4.2 – RESULTADOS RONDA 2 PROCESSO DELPHI

As questões 11, 17, 18, 19 e 23 não foram consensuais necessitando de reformulação. As primeiras quatro obtiveram a aprovação máxima de 4 peritos (4/6) e a última de 3 peritos (3/6).

Na questão 11 o termo “point measures” não foi considerado equivalente a medidas de localização. Pelo que foi reformulado para obter melhor equivalência conceptual (tabela 4.22).

TABELA 4.22 – REFORMULAÇÃO QUESTÃO 11 – RONDA 2

Ronda Anterior	Reformulação	Nº de peritos
“...medidas de localização...”	“...medidas de precisão ...”	2

Na questão 17 foi reformulada para melhor adequação de terminologia no primeiro caso e melhor construção frásica no segundo (tabela 4.23).

TABELA 4.23 – REFORMULAÇÃO QUESTÃO 17 – RONDA 2

Ronda Anterior	Reformulação	Nº de peritos
“...medida de severidade...”	“...medida de gravidade...”	2
“...de que não é possível prever, com base nas condições de partida...”	“...de que, com base nas condições de partida, não é possível prever...”	2

A questão 18 foi reformulada para simplificação e fluência de discurso nos dois últimos parâmetros e para melhor adequação semântica no primeiro. (Tabela 4.24).

TABELA 4.24 – REFORMULAÇÃO QUESTÃO 18 – RONDA 2

Ronda Anterior	Reformulação	Nº de peritos
“...o primeiro indicador...”	“...o indicador primário ...”	2
“...há mais que uma...”	“...utilizam mais do que uma...”	2
“...que pode ser usada como...”	“...como...”	1

Na questão 19 a tradução do termo “Blinding” não foi ainda consensual sugerindo-se o referido na tabela 4.25.

TABELA 4.25 – REFORMULAÇÃO QUESTÃO 19 – RONDA 2

Ronda Anterior	Reformulação	Nº de peritos
“Estar cego ao estudo...”	“Ser cego para o estudo...”	2

Na questão 23 a frase ainda não foi considerada clara, principalmente no último parâmetro (Tabela 4.26).

TABELA 4.26 – REFORMULAÇÃO QUESTÃO 23 – RONDA 2

Ronda Anterior	Reformulação	Nº de peritos
“...medida de localização...”	“...medida de precisão ...”	4
“...desvios-padrão, erros-padrão...”	“...desvios-padrão (DP’s), erros-padrão(EP’s)...”	2
“...os resultados são categóricos...”	“...os resultados são relativos a variáveis categóricas...”	2

Mais uma vez, após reformulação, procedeu-se ao início de uma nova ronda.

4.2.1.4 Ronda 3

Na ronda 3 as questões 11, 18, 19 e 23 foram consensuais (Ilustração 4.3).

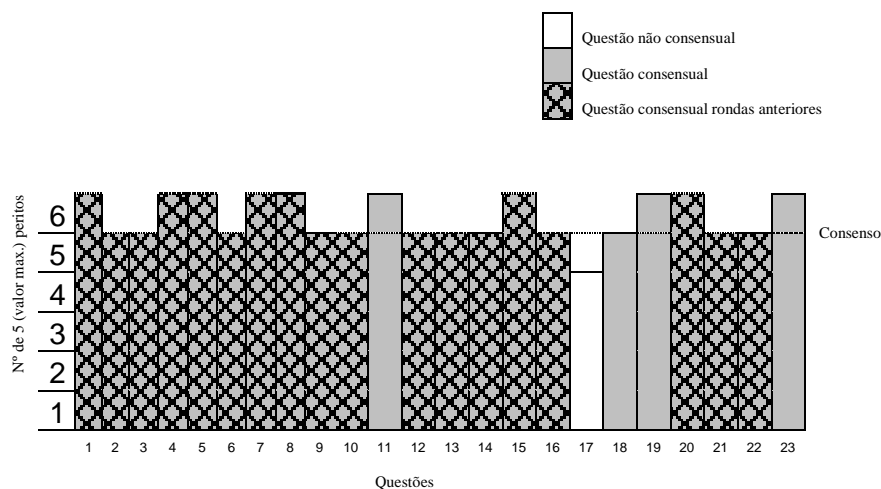


ILUSTRAÇÃO 4.3 – RESULTADOS RONDA 3 PROCESSO DELPHI

A questão 17 foi a única não consensual e manteve a mesma pontuação da ronda anterior (3/6), sendo reformulada para clarificação do discurso (Tabela 4.27).

TABELA 4.27 – REFORMULAÇÃO QUESTÃO 17 – RONDA 3

Ronda Anterior	Reformulação	Nº de peritos
“...medida de gravidade...”	“...medida da gravidade...”	2
“...com base nas condições de partida...”	“...com base nas condições de prognóstico de início...”	3
“...não é possível prever...”	“...não seja possível prever...”	2
“...é satisfeito...”	“...é atingido...”	1

Depois da reformulação procedeu-se então à ronda seguinte.

4.2.1.5 Ronda 4

Na ronda 4 a questão restante (17) foi consensual, obtendo a aprovação máxima de todos os peritos (ilustração 4.4).

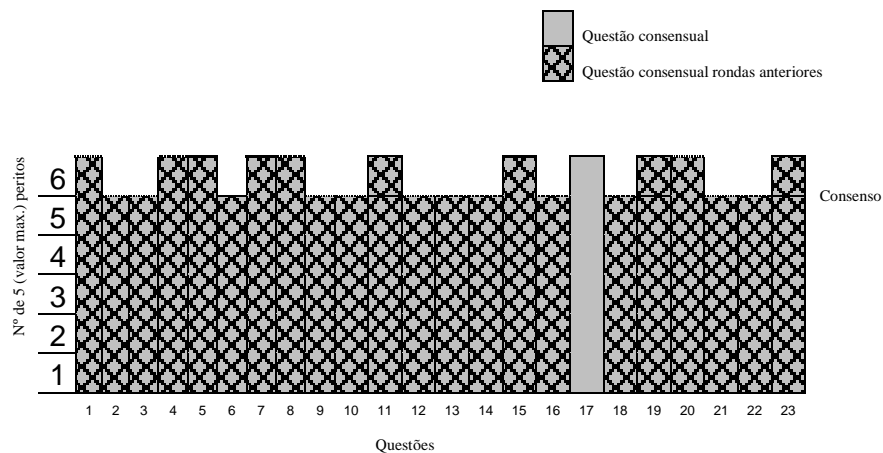


ILUSTRAÇÃO 4.4 – RESULTADOS RONDA 4 PROCESSO DELPHI

Foi então melhorada a equivalência semântica nesta fase. À parte de alterações de ordem básica (construção frásica e clarificação de discurso), o processo Delphi permitiu a definição de termos que estavam pouco descritos na literatura portuguesa, na área científica.

Por se ter considerado claro o seu objecto de avaliação da escala foi assegurada a validação de face.

4.2.2 Pré-teste

4.2.2.1 Caracterização da amostra do pré-teste

A amostra era composta por 20 fisioterapeutas (12♀ e 8♂), todos licenciados. A tabela 4.28 apresenta as principais características sócio-demográficas dos sujeitos que participaram no pré-teste.

TABELA 4.28 – DADOS SOCIO-DEMOGRÁFICOS AMOSTRA PRÉ-TESTE

Variáveis	Média (DP)	Mínimo	Máximo	Moda
Género	-	-	-	1 (feminino)
Idade	26,65 (4,671)	23	43	
Anos profissão	2,66 (3,180)	1	15	

4.2.2.2 Escala de PEDro - Concordância artigo

Conforme já descrito no capítulo da metodologia, foi sugerido pelos autores retirar artigos da *PEDro*, uma vez que já teriam sido avaliados pelos autores.

Aqui se apresentam os resultados das avaliações da amostra, comparando com o valor já previamente atribuído pelos autores.

Artigo 1 – Knobel, R., Radünz, V., & Carraro, T. E. (2005). [Use of Transcutaneous Electric Nerve Stimulation for Pain Relief in Labour: A possible way to care the child-bearer]. *Texto contexto*, 14 (2), 229-236.

Na *PEDro* este estudo estava avaliado como todos os critérios estavam presentes à exceção do critério 9 (análise de dados com intenção-de-tratamento), portanto 9.

A avaliação feita pela amostra obteve uma percentagem de concordância de 75% (Tabela 4.29) em termos gerais, sendo que o item que obteve mais concordância foi o item 10 (Comparações estatísticas inter-grupos para pelo menos um resultado-chave) com 100% e o que obteve menos foi o item 9 (já referido acima), com 40 %. Relativamente à pontuação total obteve uma percentagem de concordância de 85,56 %.

TABELA 4.29 – RESULTADOS ARTIGO 1

		Website	Pré-teste		Diferença	% de Concordância
			(Média aritmética)	(desvio-padrão)	(Pre-Web)	
Itens	1	1	0,85	0,366	-0,15	85%
	2	1	0,80	0,410	-0,20	80%
	3	1	0,75	0,444	-0,25	75%
	4	1	0,75	0,444	-0,25	75%
	5	1	0,65	0,489	-0,35	65%
	6	1	0,60	0,503	-0,40	60%
	7	1	0,80	0,410	-0,20	80%
	8	1	0,75	0,444	-0,25	75%
	9	0	0,60	0,503	+0,60	40%
	10	1	1	0,000	0,00	100%
	11	1	0,85	0,366	-0,15	85%
Total		9/10	7,70/10	2,105	-1,30	85,56%
Média						75%

Artigo 2 – Ashburn, A., Fazakarley, L., Ballinger, C., Pickering, R., McLellan, L. D., & Fitton, C. (2007). A randomised controlled trial of a home based exercise programme to reduce the risk of falling among people with Parkinson's disease. *J Neurol Neurosurg Psychiatry*, 78(7), 678-684.

Na *PEDro* este estudo estava avaliado como todos os critérios estavam presentes à exceção do critério 5 (*Cegueira* sujeitos) e 6 (*cegueira* fisioterapeutas), portanto 8.

A avaliação feita pela amostra obteve uma percentagem de concordância de 77% (Tabela 4.30) em termos gerais, sendo que o item que obteve mais concordância foram os itens 4 (semelhança de grupos a início) e 10 (Comparações estatísticas inter-grupos para pelo menos um resultado-chave) com 100% e o que obteve menos foram os itens 5 e 6 (já referidos acima), com 20 %. Relativamente à pontuação total obteve uma percentagem de concordância de 109,38% por ser a mais, sendo portanto de 90,62 %.

TABELA 4.30 – RESULTADOS ARTIGO 2

		Website	Pré-teste		Diferença	% de Concordância
			(Média aritmética)	(desvio-padrão)	(Web-pre)	
Itens	1	1	,80	,410	-0,20	80 %
	2	1	,90	,308	-0,10	90%
	3	1	,85	,366	-0,15	85 %
	4	1	1,00	,000	0	100%
	5	0	,80	,410	+0,80	20%
	6	0	,80	,410	+0,80	20%
	7	1	,90	,308	-0,10	90%
	8	1	,85	,366	-0,15	85%
	9	1	,70	,470	-0,30	70%
	10	1	1,00	,000	0	100%
	11	1	,90	,308	-0,10	90%
Total		8/10	8,75	1,209	+0,75	109,38% 90,62 %
Média						77%

Artigo 3 – Azevedo, D. C., Carvalho, S. C., Leal, E., Damasceno, S. P., & Ferreira, M. L. (2008). [Influence of the range of motion (Rom) Limitation on the Shoulder Flexibility Improvement after six-week-training]. *Revista Brasileira Medicina do Esporte*, 14(2), 119-121.

Na *PEDro* este estudo estava avaliado com uma pontuação de 6, não cumprindo os critérios 1 (elegibilidade), 3 (Distribuição cega), 5 (*Cegueira* sujeitos), 6 (*cegueira* fisioterapeutas) e 9 (análise de dados com intenção-de-tratamento).

A avaliação feita pela amostra obteve uma percentagem de concordância de 74% (Tabela 4.31) em termos gerais, sendo que os itens que obtiveram mais concordância foram os itens 4 (semelhança de grupos a início) e 8 (medição resultados-chave em mais de 85%) com 100% e o que obteve menos foi o item 9 (já referido acima), com 0 %. Relativamente á pontuação total obteve uma percentagem de concordância de 126,84% por ser a mais, sendo portanto de 73,16%.

TABELA 4.31 – RESULTADOS ARTIGO 3

		Website	Pré-teste		Diferença	% de Concordância
			(Média aritmética)	(desvio-padrão)	(Web-pre)	
Itens	1	0	0,94	0,236	+0,94	6%
	2	1	0,89	0,323	-0,11	89%
	3	0	0,50	0,514	+0,50	50%
	4	1	1,00	0,000	0	100%
	5	0	0,44	0,511	+0,44	66%
	6	0	0,44	0,511	+0,44	66%
	7	1	0,56	0,511	-0,44	56%
	8	1	1,00	0,000	0	100%
	9	0	1,00	0,000	0	0%
	10	1	0,94	0,236	-0,06	95%
	11	1	0,83	0,383	-0,17	83%
Total		6/10	7,61	1,650	+1,61	126,84% 73,16%
Média						74%

4.2.2.3 Formulário de compreensão

Relativamente ao valor esperado cumpriu-se o critério pré-definido para aprovação: em todos os itens obteve-se ≥ 4 (compreensível a claro), por média aritmética, com um desvio padrão de no máximo ± 1 (algumas dúvidas a claro). Tais valores podem-se verificar na Ilustração 4.5 e Tabela 4.32.

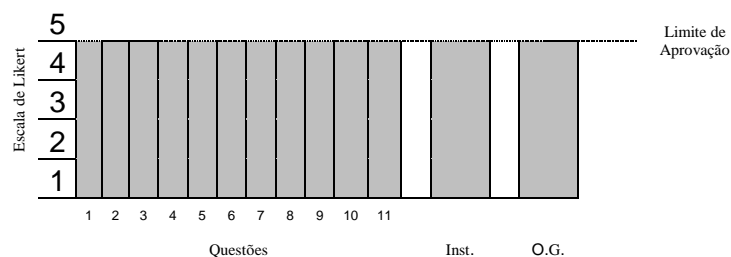


ILUSTRAÇÃO 4.5 – GRÁFICO DOS RESULTADOS FORMULÁRIO DE COMPREENSÃO

O critério 9 foi o que suscitou mais dúvidas e também o que obteve menor valor quanto à compreensão, 3,55 (Tabela 4.32). Como observamos na tabela nenhum item foi considerado incompreensível, estando todos no nível compreensível.

TABELA 4.32 – RESULTADOS FORMULÁRIO DE COMPREENSÃO

	N	Mínimo	Máximo	Média	Desvio-padrão
Item 1	20	3	5	4,10	0,968
Item 2	20	4	5	4,40	0,503
Item 3	20	3	5	4,15	0,745
Item 4	20	3	5	4,40	0,754
Item 5	20	2	5	3,85	0,875
Item 6	20	2	5	4,00	0,858
Item 7	20	3	5	4,00	0,795
Item 8	20	2	5	3,70	0,801
Item 9	20	2	5	3,55	0,826
Item 10	20	3	5	4,45	0,605
Item 11	20	3	5	4,00	0,795
Instruções	20	2	5	3,60	0,754
Opinião geral	20	3	5	4,00	0,649
Valid N (listwise)	20				

Quanto às respostas abertas o que foi significativamente abordado pela amostra foi a complexidade e especificidade da linguagem e a necessidade de outro tipo de conhecimento ou treino para o preenchimento da Escala de PEDro.

Também foi referido que sentiam falta de segurança nas opções e achavam que poderia enviesar os resultados.

O critério 9 foi o mais comentado quanto à sua compreensão e à sua relevância, facto que, pode ser explicado precisamente pela necessidade de conhecimentos mais alargados na área.

Foi também constatado um erro ortográfico no critério 9: “...que lhes tido sido atribuído inicialmente...” em vez de “...que lhes tinha sido atribuído inicialmente...”. Foi corrigido na versão final.

Foi também referido pelo pré-teste que as instruções poderiam ser um pouco mais claras e objectivas.

Tais achados vão de encontro ao estabelecido desde início, que é necessário treino específico para a aplicação da Escala de PEDro.

Verificou-se também que não se encontraram diferenças relativas ao idioma em que o artigo está redigido, quer por valores (75%, 77% e 74%), quer por referência dos participantes nas respostas abertas.

Por se ter considerado claro o objecto de avaliação da escala foi assegurada a validação de face.

4.2.2.4 Tempo de preenchimento

Por apresentar uma diversidade muito alargada no intervalo de tempo relativo ao preenchimento (Tabela 4.33) e pela dificuldade já descrita anteriormente, (apesar do valor uniforme de ± 30 minutos) este resultado não poderá ser considerado como valor padrão, a não ser para populações não treinadas.

TABELA 4.33 – RESULTADOS TEMPO DE PREENCHIMENTO

Tempo Preenchimento	N	Range	Mínimo	Máximo	Média (DP)
Artigo1	20	36	16	52	33,65 (11,554)
Artigo 2	20	42	13	55	28,35 (10,378)
Artigo 3	20	35	15	50	30,60 (9,321)

De qualquer forma como descrito no capítulo da revisão da literatura, a escala original não dispõe de valor padrão para o tempo de preenchimento.

Capítulo V

Discussão e Conclusão

Neste capítulo vai-se discutir os resultados, tendo em conta as diferentes fases e o que era esperado, retirando conclusões para a aplicação prática do estudo.

Toda a sequência deste estudo correu conforme o descrito na literatura (73), sem necessidade de repetir etapas.

5.1 Processo Delphi

Comparando com o descrito na literatura (27), o processo Delphi correu conforme esperado. Não foi necessário estudar o nível de concordância uma vez que foi definido o consenso com o valor máximo para quase todo o grupo (5/6). As quatro rondas necessárias foram todas produtivas e não repetitivas, o que muitas vezes é um risco neste método (77). O principal problema colocou-se em alguma terminologia específica que, através das sugestões, conseguiu-se encontrar um conjunto de definição de termos bastante adequado para a cultura portuguesa.

Como reforço do sucesso desse resultado está o facto da versão produzida pelo painel sofrer poucas alterações para compor a versão final.

5.2 Pré-teste

O pré-teste cumpriu os objectivos propostos (73): avaliar a compreensão da Escala de PEDro quanto à sua tradução e adaptação à língua portuguesa. O valor geral da escala de *likert* foi aceitável ($4 \pm 0,764$), estando dentro dos padrões exigidos. Pois a escala obteve o nível de compreensível em todos os itens. O nível de concordância foi de 75 %, não foi mais elevado provavelmente pela falta de treino dos participantes, que aliás foi referida pelos próprios. As dúvidas que surgiram entre os participantes relacionavam-se com a complexidade e especificidade do tema e não com a linguagem em que estava descrito. Tal achado prende-se ao facto de amostra não dominar a temática da escala.

Este trata-se então de outro objectivo cumprido, o da averiguação da necessidade de treino específico para a aplicação desta escala (13; 46; 49; 58). Tal necessidade verificou-se efectivamente. No entanto, com um nível de concordância de 75 % poder-se-á sugerir que fisioterapeutas licenciados, portanto com alguma noção de investigação, são capazes de utilizar esta escala com uma margem de erro de 25 %.

Quanto ao tempo de preenchimento, que não está descrito na literatura (53), verificou-se um intervalo médio de ± 30 minutos. Porém só se poderá sugerir que sujeitos não treinados levem esse tempo. São necessários futuros estudos para apurar o tempo de preenchimento de indivíduos treinados.

As alterações referidas foram de ordem ortográfica, não se verificando dificuldades ao nível de compreensão e relevância cultural da tradução. Verificou-se dificuldades na interpretação mas pelo motivo já explicitado.

Como já referido o idioma em que o artigo está redigido não suscitou alterações na aplicação do instrumento.

Relativamente à validade de face esta foi obtida através do painel de peritos e pré-teste, conforme descrito (27).

5.3 Conclusões

Neste trabalho foi feita a tradução e adaptação cultural de um instrumento que permite avaliar o rigor científico de ECR's e ECC's no âmbito da Fisioterapia – A Escala de PEDro (tabela 5.1).

TABELA 5.1 – ESCALA DE PEDRO

Critério	Descrição	sim/não
1.	Especificação do critério de elegibilidade	
2.	Colocação aleatória	
3.	Distribuição cega.	
4.	Semelhança de grupos ao início	
5.	Sujeitos cegos aos estudos.	
6.	Fisioterapeutas cegos ao estudo.	
7.	Avaliadores cegos ao estudo.	
8.	Medição resultados-chave em mais de 85%.	
9.	Análise de intenção de tratamento.	
10.	Análise inter-grupos para pelo menos um resultado-chave.	
11.	Medidas de precisão e variabilidade para pelo menos um resultado-chave.	

Com os resultados apresentados pode-se afirmar que esta versão é equivalente em termos semânticos, idiomáticos e conceptuais à medida original.

Conclui-se então que a Escala de PEDro tem grande utilidade na prática clínica e que foi demonstrado apresentar validade de face para a cultura portuguesa.

Mantém-se, no entanto, em aberto, a verificação das restantes propriedades psicométricas.

Capítulo VI

Considerações Finais

Neste capítulo descrevem-se as limitações deste estudo e o desenho de estudo de futuros projectos inerentes ao mesmo.

6.1 Limitações

O facto de não se ter procedido à validação constitui uma limitação para o estudo. Efectivamente ainda é necessário perceber se as características psicométricas estão mantidas. O facto de não existir uma escala com o mesmo objecto de avaliação que sirva de “golden standart” para estudar a validade de constructo e a necessidade da aplicação da Escala de PEDro ser feita por um avaliador treinado, também constituiu uma limitação para o estudo.

No entanto estas limitações serão contornadas pela execução de projectos abaixo descritos.

6.2 Futuros Projectos

O *PEDro steering Group*, mais especificamente Anne Moseley e Leonardo Costa, acompanharam de perto o desenvolvimento deste trabalho. Neste momento foram também desenvolvidas as versões portuguesa (Brasil), alemã, chinesa e francesa (48).

Foi proposto uma validação conjunta, isto é, com o mesmo desenho de estudo para todas as versões em colaboração com o *PEDro steering Group*.

O desenho de estudo sugerido na altura é o representado na Ilustração 6.1. Pretende-se com este desenho de estudo avaliar a fiabilidade (consistência interna, fiabilidade intra e inter-examinador). Posteriormente avaliar-se-á a validade.

Relativamente à análise estatística, a reprodutibilidade será calculada da seguinte forma:

- A fiabilidade para cada item individual da Escala de PEDro será avaliada utilizando o coeficiente de Kappa;
- A fiabilidade da pontuação total da Escala de PEDro será avaliada utilizando o Coeficiente de Correlação Intraclass tipo 1,1 ($ICC_{1,1}$);
- Concordância entre as duas versões será avaliada utilizando o Standard Error of the Measurement ($SEM_{consistency}$). O SEM será calculado dividindo o desvio-padrão da diferença média entre as

duas pontuações por $\sqrt{2}$ (i.e. $SEM = SD_{\text{differences}} / \sqrt{2}$). O SEM reflecte o erro do próprio instrumento.

- Fiabilidade Inter-examinador e concordância serão calculadas de forma similar (Kappa, ICC_{1,1} e SEM).
- A consistência interna será avaliada utilizando o Coeficiente Alfa de Cronbach.

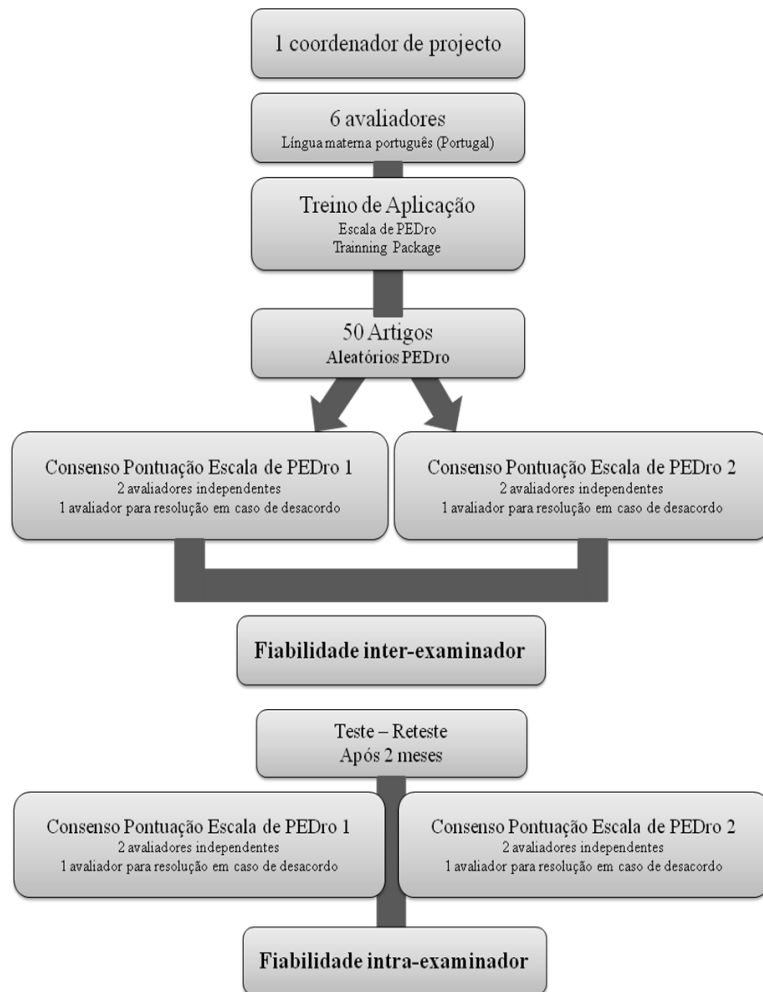


ILUSTRAÇÃO 6.1 – DESENHO DE ESTUDO PARA AVERIGUAÇÃO DE REPRODUTIBILIDADE

Será portanto o próximo passo para dar continuidade ao trabalho que compõe esta dissertação.

Bibliografia

Referências Bibliográficas

1. Akobeng, A. K. (2005). Understanding systematic reviews and meta-analysis. *Arch Dis Child*, 90(8), 845-848.
2. Altman, D. G., & Moher, D. (1999-2011). CONSORT Organization. 2010, from <http://www.consort-statement.org>
3. Armstrong, J. C. (2003-2005). Delphi Decision Aid. 2008, from <http://armstrong.wharton.upenn.edu/delphi2/>
4. Ashburn, A., Fazakarley, L., Ballinger, C., Pickering, R., McLellan, L. D., & Fitton, C. (2007). A randomised controlled trial of a home based exercise programme to reduce the risk of falling among people with Parkinson's disease. *J Neurol Neurosurg Psychiatry*, 78(7), 678-684.
5. Azevedo, D. C., Carvalho, S. C., Leal, E., Damasceno, S. P., & Ferreira, M. L. (2008). [Influence of the range of motion (Rom) Limitation on the Shoulder Flexibility Improvement after six-week-training]. *Revista Brasileira Medicina do Esporte*, 14(2), 119-121.
6. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191.
7. Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*, 276(8), 637-639.
8. Bhogal, S. K., Teasell, R. W., Foley, N. C., & Speechley, M. R. (2005). The PEDro scale provides a more comprehensive measure of methodological quality than the Jadad scale in stroke rehabilitation literature. *J Clin Epidemiol*, 58(7), 668-673.
9. Brand, R. A. (2009). Standards of reporting: the CONSORT, QUORUM, and STROBE guidelines. *Clin Orthop Relat Res*, 467(6), 1393-1394.
10. Clarke, M. (2000). The QUORUM statement. *Lancet*, 355(9205), 756-757.
11. de Morton, N. A. (2009). The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. *Aust J Physiother*, 55(2), 129-133.
12. Dias, R. C., & Dias, J. M. D. (2006). Evidence-Based Practice: a Methodology for a best Physical Therapy Practice. *Fisioterapia em Movimento*, 19(1), 11-16.
13. Elkins, M. R., Herbert, R. D., Moseley, A. M., Sherrington, C., & Maher, C. (2010). Rating the quality of trials in systematic reviews of physical therapy interventions. *Cardiopulm Phys Ther J*, 21(3), 20-26.

14. Elwyn, G., O'Connor, A., Stacey, D., Volk, R., Edwards, A., Coulter, A., et al. (2006). Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *Bmj*, 333(7565), 417.
15. Finger, M. E., Cieza, A., Stoll, J., Stucki, G., & Huber, E. O. (2006). Identification of intervention categories for physical therapy, based on the international classification of functioning, disability and health: a Delphi exercise. *Phys Ther*, 86(9), 1203-1220.
16. Foley, N. C., Bhogal, S. K., Teasell, R. W., Bureau, Y., & Speechley, M. R. (2006). Estimates of quality and reliability with the physiotherapy evidence-based database scale to assess the methodology of randomized controlled trials of pharmacological and nonpharmacological interventions. *Phys Ther*, 86(6), 817-824.
17. Foley, R. N. (2009). Analysis of randomized controlled clinical trials. *Methods Mol Biol*, 473, 113-126.
18. Fortin, M.-F. (2009). *Fundamentos e Etapas no Processo de Investigação: Lusodidacta*.
19. Garg, A. X., Hackam, D., & Tonelli, M. (2008). Systematic review and meta-analysis: when one study is just not enough. *Clin J Am Soc Nephrol*, 3(1), 253-260.
20. Green, S. (2005). Systematic reviews and meta-analysis. *Singapore Med J*, 46(6), 270-273; quiz 274.
21. Guillemin, F. (1995). Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol*, 24(2), 61-63.
22. Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*, 46(12), 1417-1432.
23. Harbour, R., & Miller, J. (2001). A new system for grading recommendations in evidence based guidelines. *BMJ*, 323(7308), 334-336.
24. Herbert, R., Jamtvedt, G., & Hagen, K. B. (2005). *Practical evidence-based physiotherapy* (2nd ed.). Edinburgh: Elsevier/Butterworth Heinemann.
25. Herbert, R., Moseley, A., & Sherrington, C. (1998). PEDro: a database of randomised controlled trials in physiotherapy. *Health Inf Manag*, 28(4), 186-188.
26. Herbert, R., Sherrington, C., Moseley, A., & Maher, C. (2000). PEDro. *Man Ther*, 5(1), 49.
27. Hicks, C. M. (2006). *Research Methods for Clinical Therapists: Lusociência*.
28. Jewell, D. V. (2008). *Guide to evidence-based physical therapy practice*. Sudbury, Mass.: Jones and Bartlett.
29. Juni, P., Holenstein, F., Sterne, J., Bartlett, C., & Egger, M. (2002). Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol*, 31(1), 115-123.
30. Kaste, M. (2006). Evidence, education and practice. *Cerebrovasc Dis*, 22(5-6), 342-349.
31. Knobel, R., Radünz, V., & Carraro, T. E. (2005). [Use of Transcutaneous Electric Nerve Stimulation for Pain Relief in Labour: A possible way to care the child-bearer]. *Texto contexto*, 14(2), 229-236.

32. Lessard-Hébert, M., Goyette, G., & Boutin, G. (2010). *Investigação Qualitativa: Fundamentos e Práticas: Epistemologia e Sociedade* - Instituto Piaget.
33. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, 339, b2700.
34. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., et al. (2009). PRISMA Organization. 2010, from <http://www.prisma-statement.org>
35. Macedo, L. G., Elkins, M. R., Maher, C. G., Moseley, A. M., Herbert, R. D., & Sherrington, C. (2010). There was evidence of convergent and construct validity of Physiotherapy Evidence Database quality scale for physiotherapy trials. *J Clin Epidemiol*, 63(8), 920-925.
36. Maher, C. G., Moseley, A. M., Sherrington, C., Elkins, M. R., & Herbert, R. D. (2008). A description of the trials, reviews, and practice guidelines indexed in the PEDro database. *Phys Ther*, 88(9), 1068-1077.
37. Maher, C. G., Sherrington, C., Elkins, M., Herbert, R. D., & Moseley, A. M. (2004). Challenges for evidence-based physical therapy: accessing and interpreting high-quality evidence on therapy. *Phys Ther*, 84(7), 644-654.
38. Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*, 83(8), 713-721.
39. Malay, D. S. (2007). Levels of clinical evidence. *J Foot Ankle Surg*, 46(2), 63-64.
40. Marques, A. P., & Peccin, M. S. (2005). Pesquisa em fisioterapia: a prática baseada em evidências e modelos de estudos. *FISIOTERAPIA e PESQUISA*, 11 (1), 43-48.
41. McKenna, S. P., & Doward, L. C. (2005). The translation and cultural adaptation of patient-reported outcome measures. *Value Health*, 8(2), 89-91.
42. Mills, E. J., Chan, A. W., Wu, P., Vail, A., Guyatt, G. H., & Altman, D. G. (2009). Design, analysis, and presentation of crossover trials. *Trials*, 10, 27.
43. Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet*, 354(9193), 1896-1900.
44. Moher, D., Cook, D. J., Jadad, A. R., Tugwell, P., Moher, M., Jones, A., et al. (1999). Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assess*, 3(12), i-iv, 1-98.
45. Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *J Am Podiatr Med Assoc*, 91(8), 437-442.

46. Moseley, A., Maher C, Herbert RD and Sherrington C. (1999). *Reliability of a scale for measuring the methodological quality of clinical trials*. Paper presented at the Proceedings of the VIIth Cochrane Colloquium., Rome.
47. Moseley, A. M., Elkins, M. R., Herbert, R. D., Maher, C. G., & Sherrington, C. (2009). Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol*.
48. Moseley, A. M., Herbert, R., Maher, C. G., Sherrington, C., & Elkins, M. R. (1999-2011). Physiotherapy Evidence Database. 2007, from <http://www.pedro.org.au/>
49. Moseley, A. M., Herbert, R., Maher, C. G., Sherrington, C., & Elkins, M. R. (2008). PEDro scale can only rate what papers report. *Aust J Physiother*, 54(4), 288.
50. Moseley, A. M., Herbert, R. D., Maher, C. G., Sherrington, C., & Elkins, M. R. (2011). Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. *J Clin Epidemiol*, 64(6), 594-601.
51. Moseley, A. M., Herbert, R. D., Sherrington, C., & Maher, C. G. (2002). Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). *Aust J Physiother*, 48(1), 43-49.
52. Needleman, I. (2000). Is this good research? Look for CONSORT and QUORUM. *Evidence-Based Dentistry*, 2, 61-62.
53. Olivo, S. A., Macedo, L. G., Gadotti, I. C., Fuentes, J., Stanton, T., & Magee, D. J. (2008). Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther*, 88(2), 156-175.
54. Pereira, A. C. (2010). Serviço de Bioestatística e Informática Médica Faculdade de Medicina da Universidade do Porto. 2010, from <http://stat2.med.up.pt/>
55. Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research : applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
56. Roberts, C., & Sibbald, B. (1998). Understanding controlled trials. Randomising groups of patients. *BMJ*, 316(7148), 1898-1900.
57. Roberts, C., & Torgerson, D. J. (1999). Understanding controlled trials: baseline imbalance in randomised controlled trials. *BMJ*, 319(7203), 185.
58. Salter, K. L., Teasell, R. W., Foley, N. C., & Jutai, J. W. (2007). Outcome assessment in randomized controlled trials of stroke rehabilitation. *Am J Phys Med Rehabil*, 86(12), 1007-1012.
59. Sampaio, R. F., & Mancini, M. C. (2007). Systematic Review Studies: A Guide for Careful Synthesis of Scientific Evidence. *Revista Brasileira de Fisioterapia*, 11(1), 77-82.
60. Schlosser, R. W. (2007). Appraising the quality of Systematic Reviews. *Focus*, 17(1), 2-8.
61. Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials*, 11, 32.

62. Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, 273(5), 408-412.
63. Sherrington, C., Moseley, A. M., Herbert, R. D., Elkins, M. R., & Maher, C. G. (2010). Ten years of evidence to guide physiotherapy interventions: Physiotherapy Evidence Database (PEDro). *Br J Sports Med*, 44(12), 836-837.
64. Sibbald, B., & Roberts, C. (1998). Understanding controlled trials. Crossover trials. *BMJ*, 316(7146), 1719.
65. Sibbald, B., & Roland, M. (1998). Understanding controlled trials. Why are randomised controlled trials important? *BMJ*, 316(7126), 201.
66. Smith, E., Jennings, P., McDonald, S., MacPherson, C., O'Brien, T., & Archer, F. (2007). The Cochrane Library as a resource for evidence on out-of-hospital health care interventions. *Ann Emerg Med*, 49(3), 344-350.
67. Thordarson, D. B. (2008). Editorial: levels of evidence, evidence-based medicine, and Foot & Ankle International. *Foot Ankle Int*, 29(9), 881-882.
68. Vaarbakken, K., Ljunggren, A. E., & Hendriks, E. J. (2008). The importance of the new CONSORT Statement for clinicians. *Aust J Physiother*, 54(3), 155-156.
69. van Everdingen, J. J. (2006). [Levels of evidence value are useful in the evaluation of studies and results]. *Ned Tijdschr Geneesk*, 150(45), 2484.
70. van Tulder, M., Furlan, A., Bombardier, C., & Bouter, L. (2003). Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976)*, 28(12), 1290-1299.
71. Verhagen, A. P., de Vet, H. C., de Bie, R. A., Boers, M., & van den Brandt, P. A. (2001). The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol*, 54(7), 651-654.
72. Verhagen, A. P., de Vet, H. C., de Bie, R. A., Kessels, A. G., Boers, M., Bouter, L. M., et al. (1998). The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*, 51(12), 1235-1241.
73. Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*, 8(2), 94-104.
74. Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Juni, P., Altman, D. G., et al. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*, 336(7644), 601-605.
75. Wright, J. G., Swiontkowski, M., & Heckman, J. D. (2006). Levels of evidence. *J Bone Joint Surg Br*, 88(9), 1264.
76. Wright, J. G., Swiontkowski, M. F., & Heckman, J. D. (2003). Introducing levels of evidence to the journal. *J Bone Joint Surg Am*, 85-A(1), 1-3.

77. Yousuf, M. I. (2007). The Delphi Technique. *Essays in Education*, 20, 8089.

Apêndices

Apêndice I – Pedido de autorização a autores da PEDro Scale

A autorização foi pedida via e-mail:

Dear Colleagues

First of all I want to congratulate for the exceptional work on the organization of Physiotherapy Evidence Database. I think it is a very good tool for us physiotherapists.

I am Cláudia Costa, a Master in Physiotherapy Sciences student, on Faculdade de Motricidade Humana (LISBON) and I was thinking to translate and adapt the PEDro scale for portuguese (Portugal) language.

I want to know who I ask permission for this assignment.

Please accept my sincere thanks for your time and consideration of my request.

I look forward to hearing from you.

Sincerely,

Cláudia Costa

Dear Anne Mosely,

My name is Jan Cabri. I am an associate professor at the Technical University of Lisbon, Faculty of Human Movement Studies, where I direct the Research Center for Physical Therapy.

I am writing on behalf of Claudia Costa, one of my Masters students who wants to translate the PEDro scale Portuguese and adapt the scale to the Portuguese reality.

Please, find for this purpose, the procedures we want to implement in order to have the PEDro scale adapted to the Portuguese culture, as attachment to this message. May I remind you that this translation is different from the Brazilian one, since there are some major differences in both culture and language.

I do hope we may come to a collaboration and that we may be able to publish the manuscript as soon as we get your approval and collaboration (it is my idea to prepare a publication in mutual cooperation, should you agree, of course). Hope to hear from you soon.

Best regards,

Jan Cabri, PT, PhD

Jan MH Cabri
Universidade Técnica de Lisboa
Faculdade de Motricidade Humana
Estrada da Costa
P-1495-688 Cruz Quebrada, Lisboa
Portugal
Tel + 351 21 4149164
Fax + 351 21 4144712

Apêndice II – Relatório enviado aos Autores “Draft Protocol”

Translation, cultural adaptation and validation of PEDro scale into Portuguese (Portugal) language

Evidence-based practice has become one of the greatest concerns for both patients and health-care professionals.

The need of research leads to improper methods and not reliable results. To be able to filter the information it is necessary a reliable and structured evaluation instrument.

PEDro Scale has the goal to qualify and evaluate the methodology of randomised control trials and clinical controlled trials.

Methods

Translation and cultural adaptation

The Portuguese translation will follow the sequential forward and backward translation approach according to Beaton et al (2000) (6) and Wild et al (2005) (73) (figure 1):

- 1) Preparation – Authors authorization and involvement. Preparation of the methodology;
- 2) Forward Translation - translation of PEDro scale original language version into Portuguese (Portugal) by translator 1 (a bilingual Portuguese native-speaker);
- 3) Reconciliation - comparing and merging more than one expert opinion of the forward translation (first version). This will be done by a web based Delphi Consensus Process: a panel of 6 experts evaluated and criticised the first translation. The evaluation will be done through an ordinary scale (1 – do not agree to 5 – totally agree with the translation) and had to be consensual (5 or more experts' ratings of 5 points). Then the investigator will perform the corrections and proceed to another round until all the questions are consensual. In the end when all questions were consensual about semantic equivalency. The Delphi process will produce the second version of the forward translation.
- 4) Back Translation - translation of the Portuguese version back into the original language by translator 2 (a bilingual English native speaker);
- 5) Back Translation Review and Report- comparison of the back-translated version with the original to highlight and investigate discrepancies between the original and the reconciled translation by the authors of the PEDro Scale. The back-translated version has to be semantically equivalent to the PEDro Scale original version;
- 6) Pre-test Cognitive Debriefing - testing the instrument on a small group in order to test alternative wording and to check understanding,

interpretation, and cultural relevance of the translation. This second version will be applied to 10-30 physiotherapists for 3 published RCTs (2 in Portuguese and 1 in English - appendix 1). The 3 RCTs reports were part of the PEDro database. A comprehension survey with an ordinal scale (1 - unclear to 5 - clear) will be filled by them also;

- 7) Review of Cognitive Debriefing Results and Finalization - comparison of the participants' interpretation of the second version to highlight and amend discrepancies, and to compose the third and final Portuguese version;
- 8) Proofreading - final review of the translation to highlight and correct any typographic, grammatical or other errors;
- 9) Final version and Report – the final version of Escala de PEDro report written at the end of the process documenting the development of each translation. When this step is finished we can proceed to the validation process.

Transcultural Validation

Transcultural validity testing (figure 1) will be assessed in a cross-sectional study. We will apply the final version of Escala de PEDro to 100 physiotherapists at list (10*10 items to insure *external validity*) for the 3 RCTs reports (Appendix 1) in two moments in time (test – retest). The second moment will be one month after. *External validity* refers to the approximate truth of conclusions that involves generalizations.

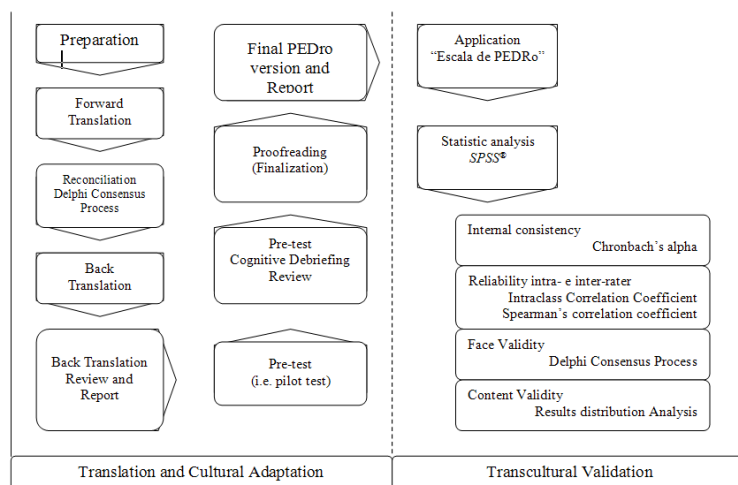


Figure 1 - Study design flowchart

Reliability

Internal consistency - homogeneity

The Chronbach's alpha quantified the level of the relationship between different items within the questionnaire (internal consistency, homogeneity, item to total correlation), and thus determined how well the instrument, as a single entity, measured the individual properties (0.00 = none, 1.00 = perfect homogeneity). Chronbach's alpha of the single items represents the homogeneity between the test-retest results.

Test-retest reliability – reproducibility - intra-rater reliability

The Intraclass Correlation Coefficient (ICC) will be calculated for two observation time points of one item, under constant conditions. An ICC value of 0.00 indicated 'no reliability', >0.75 was defined as 'good' and 1.00 as 'perfect'. It will be also determined by using Spearman's correlation coefficient.

Inter-rater reliability

The Intraclass Correlation Coefficient (ICC) will be calculated for two observer's points of one item, under constant conditions. An ICC value of 0.00 indicated 'no reliability', >0.75 was defined as 'good' and 1.00 as 'perfect'. It will be also determined by using Spearman's correlation coefficient.

Validity

Face Validity

The "Escala de PEDro" will be sent to a panel of experts to determine face validity through another Delphi Consensus process. The experts were asked to make remarks or comments on the plausibility of the questions, the comprehensiveness, and the relevance of a scale ranging from 1 to 5 (very relevant to irrelevant).

Content Validity

The final version of Escala de PEDro should demonstrate content validity if it had few missing responses, use the full range of scores with little skew (the amount of asymmetry in the distribution of scores), and have few ceiling (best possible score) or floor (poorest possible score) effects.

Note:

Criterion Validity and Construct Validity

The criterion validity is the ability of the measurement instrument or measurement method to measure what it claims to measure (defined as the gold standard, i.e. the perfect measurement of a characteristic). A tested instrument for physiotherapy RCT methodology assessment that could serve as the gold standard for the criterion validity has not been reported in the literature. In addition, there is no instrument in Portugal with a comparable content or construct that would serve as a suitable reference for the definition of construct validity. So criterion and construct validity will not be studied.

Statistical Analysis

Descriptive statistical methods will be used to characterise the participants and to test for possible floor and ceiling effects. The test-retest reliability will be quantified, as usual, by means of the ICC, resulting from analysis of variance, and internal consistency will be quantified by means of Chronbach's alpha. For every ICC coefficient, the corresponding 95% confidence interval is also stated as a measure of precision. The other correlations relevant to validity measurement will be calculated according to Spearman's non-parametric rank method. The level of significance will be defined as $P < 5\%$ (type I error). Data analysis will be performed using Statistical Package for the Social Sciences Version 17.0 for PC (SPSS)

References

1. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25(24):3186-91.
2. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8(2):94-104.

Appendix 1

RCT report 1 – Azevedo, D. C., De Carvalho, S. C., Leal, E. S., Damasceno, S. P. & Ferreira, M. L. (Influence of the range of motion (ROM) limitation on the shoulder flexibility improvement after a six-week training) [Portuguese] *Revista Brasileira de Medicina do Esporte* 2008 Mar-Apr;14(2):119-121

RCT report 2 – Knobel, R., Radunz, V. & Carraro, T. E. (Use of transcutaneous electric nerve stimulation for pain relief in labour: a possible way to care the child-bearer) [Portuguese] *Texto & Contexto-Enfermagem* 2005;14(2):229-236

RCT report 3 – Ashburn, A., Fazakarley, L., Ballinger, C., Pickering, R., McLellan, L. D. & Fitton, C. A. Randomised controlled trial of a home based exercise programme to reduce the risk of falling among people with Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2007;78(7):678-84

Apêndice III – Tradução PEDro Scale de inglês para Português versão 1

Escala PEDro

-
- | | |
|--|---|
| 1. Os critérios de elegibilidade foram especificados? | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 2. Os sujeitos foram aleatoriamente distribuídos por grupos (num estudo transversal, os sujeitos foram aleatoriamente ordenados para receber tratamento) | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 3. A distribuição pelos grupos foi sigilosa | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 4. Os grupos eram, à partida, semelhantes no que diz respeito aos mais importantes indicadores de prognóstico | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 5. Estudo cego em relação aos sujeitos | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 6. Estudo cego em relação aos terapeutas que administraram a terapia | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 7. Estudo cego em relação aos avaliadores que mediram pelo menos um resultado-chave | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 8. Medições de pelo menos um resultado-chave foram efectuadas para mais de 85% dos sujeitos inicialmente distribuídos pelos grupos | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 9. Todos os sujeitos para os quais se apresentam medições de resultados receberam o tratamento ou a condição de controlo como previsto ou, quando não foi esse o caso, fez-se a análise dos dados para pelo menos um dos resultados-chave por “intenção de tratamento” | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 10. Os resultados das comparações estatísticas inter-grupais foram descritos para pelo menos um resultado-chave | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
| 11. O estudo apresenta pontos de medição e medições de variação para pelo menos um resultado-chave | não <input type="checkbox"/> sim <input type="checkbox"/> onde: |
-

A escala PEDro baseia-se na lista de Delphi, desenvolvida por Verhagen e colegas no Departamento de Epidemiologia, da Universidade de Maastricht (Verhagen AP et al (1988). “The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus”. *Journal of Clinical Epidemiology*, 51(12):1235-41). A lista, na sua maior parte, baseia-se num “consenso de peritos” e não em dados empíricos. Incluíram-se na escala de PEDro dois itens adicionais, que não constavam da lista de Delphi (os itens 8 e 10 da escala de PEDro). À medida que forem disponibilizados mais dados empíricos, tornar-se-á possível ponderar os itens da escala de forma a que a pontuação obtida a partir da aplicação da escala PEDro reflecta a importância de cada um dos itens da escala.

O objectivo da escala PEDro consiste em auxiliar os utilizadores da base de dados PEDro a identificar rapidamente quais dos ensaios clínicos randomizados, conhecidos ou suspeitos, (ou seja, ECR ou ECC) arquivados na base de dados PEDro poderão ter validade interna (critérios 2-9), e poderão conter suficiente informação estatística para que os seus resultados possam ser interpretados (critérios 10-11). Um critério adicional (critério 1) que relaciona a validade externa (ou “potencial de generalização” ou “aplicabilidade” do ensaio) foi conservado

para que a lista de Delphi esteja completa, mas este critério não será usado para calcular a pontuação PEDro apresentada no endereço PEDro na internet.

A escala PEDro não deverá ser usada como uma medida da “validade” das conclusões de um estudo. Advertimos, muito especialmente, os utilizadores da escala PEDro de que estudos que revelem efeitos significativos do tratamento e que obtenham pontuação elevada na escala PEDro não fornecem, necessariamente, evidência de que o tratamento tenha sido clinicamente útil. Adicionalmente, importa saber se o efeito do tratamento foi suficientemente expressivo para poder ser considerado clinicamente justificável, se os efeitos positivos do mesmo suplantam os negativos, e também aferir a relação de custo-eficácia do tratamento. A escala não deve ser utilizada para comparar a “qualidade” de ensaios realizados em diferentes áreas de terapia, porque nalgumas áreas da prática da fisioterapia não é possível satisfazer todos os itens da escala.

Indicações para a administração da escala PEDro

Todos os

critérios

Só se atribui pontuação quando um critério foi inequivocamente satisfeito. Se numa leitura literal do relatório do ensaio for possível considerar que o critério não foi satisfeito, esse critério não deve receber pontuação.

Critério 1

Este critério pode considerar-se satisfeito quando o relatório descreve as origens dos sujeitos e a lista de critérios utilizados para determinar quais os sujeitos elegíveis para participar no estudo.

Critério 2

Considera-se que num determinado estudo houve distribuição aleatória se o relatório refere que a distribuição dos sujeitos foi aleatória. O método de aleatoriedade não precisa de ser explicitado. Sorteios tais como lançamento de dados ou moeda ao ar devem considerar-se procedimentos de distribuição aleatória. Procedimentos de distribuição quase-aleatória tais como os que se efectuam a partir do número de registo no hospital, da data de nascimento, ou de alternância, não satisfazem este critério.

Critério 3

Distribuição sigilosa significa que a pessoa que determinou a elegibilidade do sujeito para participar no ensaio desconhecia, no momento da decisão, o grupo a que o sujeito iria pertencer. Deve atribuir-se um ponto a este critério, mesmo que não se diga que a distribuição foi sigilosa, quando o relatório refere que a distribuição foi feita a partir de envelopes fechados ou que a distribuição implicou o contacto com o responsável pela distribuição dos sujeitos por grupos, e este último não estava implicado no ensaio.

- Critério 4 No mínimo, nos estudos de intervenções terapêuticas, o relatório deve descrever pelo menos uma medida do estado da condição a ser tratada e pelo menos uma (diferente) medida de resultado-chave que caracteriza o ponto de partida. O avaliador deve assegurar-se de que não é possível prever, com base nas condições de partida, diferenças de resultados para os diversos grupos.
- Critérios 4, 7-11 Resultados-chave são resultados que fornecem o primeiro indicador da eficácia (ou falta de eficácia) da terapia. Na maioria dos estudos, há mais de uma variável que pode ser usada como indicador de eficácia.
- Critérios 5-7 *Estudo Cego em relação ao sujeito, terapeuta ou avaliador* significa que a pessoa em questão (sujeito, terapeuta ou avaliador) não conhece qual o grupo em que o sujeito é integrado. O estudo só é cego em relação aos sujeitos e aos terapeutas se for possível esperar-se que os mesmos sejam incapazes de distinguir entre os tratamentos aplicados aos diferentes grupos. Nos ensaios em que os resultados-chave são relatados pelo próprio (por exemplo, escala visual análoga, registo diário da dor), o estudo é considerado cego em relação ao avaliador se o for em relação ao sujeito.
- Critério 8 Este critério só se considera satisfeito se o relatório referir explicitamente tanto o número de sujeitos inicialmente integrados nos grupos como o número de sujeitos para os quais se obtiveram medidas de resultados-chave. Nos ensaios em que os resultados são medidos em diferentes momentos, um resultado-chave tem de ter sido medido em mais de 85% dos sujeitos num destes momentos.
- Critério 9 Uma análise de *intenção de tratamento* significa que, quando os sujeitos não receberam tratamento (ou a condição de controlo) tal como previsto, e quando se encontram disponíveis medidas de resultados, a análise foi efectuada como se os sujeitos tivessem recebido o tratamento (ou a condição de controlo) que lhes tido sido atribuído inicialmente. Este critério é satisfeito, mesmo que não haja menção de análise por intenção de tratamento, se o relatório referir explicitamente que todos os sujeitos receberam o tratamento ou condição de controlo previstos.
- Critério 10 Uma *comparação inter-grupal* estatística implica uma comparação estatística de um grupo com outro. Conforme o desenho do estudo, isto pode implicar uma comparação de dois ou mais tratamentos, ou a comparação do tratamento com a condição de controlo. A análise pode ser uma simples comparação dos resultados medidos após a administração do tratamento, ou a comparação da mudança num grupo em relação à mudança no outro (quando se usou uma análise factorial de

variância para analisar os dados, esta última é frequentemente descrita como interação grupo x tempo). A comparação pode assumir a forma de uma estimativa (por exemplo, a diferença média ou a diferença mediana, ou uma diferença nas proporções, ou um número necessário para tratar, ou um risco relativo ou um rácio de risco) e respectivo intervalo de confiança.

Critério 11

Um *ponto de medida* é uma medida da dimensão do efeito do tratamento. O efeito do tratamento pode ser descrito como uma diferença nos resultados do grupo, ou como o resultado em todos os (ou em cada um dos) grupos. Medidas de variabilidade incluem desvios-padrão, erros-padrão, intervalos de confiança, amplitudes interquartis (ou outras amplitudes de quantis), e amplitudes de variação. Os pontos de medida e/ou as medidas de variabilidade podem ser apresentados graficamente (por exemplo, os DP podem ser apresentados como barras de erro numa figura) desde que aquilo que é representado seja inequivocamente identificável (por exemplo, desde que fique claro se as barras de erro representam DP ou EP). Quando os resultados são categóricos, considera-se que este critério foi satisfeito se o número de sujeitos em cada categoria é dada para cada grupo.

Apêndice IV – Rondas Processo Delphi

Processo Delphi

Ronda 1

Nº	Pergunta	Pontuação	
		Moda	Val 5
1	Concorda com a tradução do item 1 da escala de PEdro "Os critérios de elegibilidade foram especificados?" do original "eligibility criteria were specified." ?	5	6/6
2	Concorda com a tradução do item 2 da escala de PEdro "Os sujeitos foram aleatoriamente distribuídos por grupos (num estudo transversal, os sujeitos foram aleatoriamente ordenados para receber tratamento)" do original "subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)." ?	3	2/6
3	Concorda com a tradução do item 3 da escala de PEdro "A distribuição pelos grupos foi sigilosa" do original "allocation was concealed."?	5	3/6
4	Concorda com a tradução do item 4 da escala de PEdro "Os grupos eram, à partida, semelhantes no que diz respeito aos mais importantes indicadores de prognóstico" do original "the groups were similar at baseline regarding the most important prognostic indicators." ?	5	4/6
5	Concorda com a tradução do item 5 da escala de PEdro "Estudo cego em relação aos sujeitos" do original "there was blinding of all subjects."?	4	2/6
6	Concorda com a tradução do item 6 da escala de PEdro "Estudo cego em relação aos terapeutas que administraram a terapia" do original "there was blinding of all therapists who administered the therapy."?	3	1/6
7	Concorda com a tradução do item 7 da escala de PEdro "Estudo cego em relação aos avaliadores que mediram pelo menos um resultado-chave" do original "there was blinding of all assessors who measured at least one key outcome."?	3	2/6
8	Concorda com a tradução do item 8 da escala de PEdro "Medições de pelo menos um resultado-chave foram efectuadas para mais de 85% dos sujeitos inicialmente distribuídos pelos grupos" do original "measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups."?	3	3/6
9	Concorda com a tradução do item 9 da escala de PEdro "Todos os sujeitos para os quais se apresentam medições de resultados receberam o tratamento ou a condição de controlo como previsto ou, quando não foi esse o caso, fez-se a análise dos dados para pelo menos um dos resultados-chave por "intenção de tratamento"?" do original "all subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by 'intention to treat'."?	5	4/6
10	Concorda com a tradução do item 10 da escala de PEdro "Os resultados das comparações estatísticas inter-grupos foram descritos para pelo menos um resultado-chave" do original "the results of between-group statistical comparisons are reported for at least one key outcome."?	5	5/6
11	Concorda com a tradução do item 11 da escala de PEdro "O estudo apresenta pontos de medição e medições de variação para pelo menos um resultado-chave" do original "the study provides both point measures and measures of variability for at least one key outcome."?	5	4/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
12	<p>Concorda com a tradução da introdução da escala de PEDro "A escala PEDro baseia-se na lista de Delphi, desenvolvida por Verhagen e colegas no Departamento de Epidemiologia, da Universidade de Maastricht (Verhagen AP et al (1988). "The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus". Journal of Clinical Epidemiology, 51(12):1235-41). A lista, na sua maior parte, baseia-se num "consenso de peritos" e não em dados empíricos. Incluíram-se na escala de PEDro dois itens adicionais, que não constavam da lista de Delphi (os itens 8 e 10 da escala de PEDro). A medida que forem disponibilizados mais dados empíricos, tornar-se-á possível ponderar os itens da escala de forma a que a pontuação obtida a partir da aplicação da escala PEDro reflecta a importância de cada um dos itens da escala. O objectivo da escala PEDro consiste em auxiliar os utilizadores da base de dados PEDro a identificar rapidamente quais dos ensaios clínicos randomizados, conhecidos ou suspeitos, (ou seja, ECR ou ECC) arquivados na base de dados PEDro poderão ter validade interna (critérios 2-9), e poderão conter suficiente informação estatística para que os seus resultados possam ser interpretados (critérios 10-11). Um critério adicional (critério 1) que relaciona a validade externa (ou "potencial de generalização" ou "aplicabilidade" do ensaio) foi conservado para que a lista de Delphi esteja completa, mas este critério não será usado para calcular a pontuação PEDro apresentada no endereço PEDro na internet. A escala PEDro não deverá ser usada como uma medida da "validade" das conclusões de um estudo. Advertimos, muito especialmente, os utilizadores da escala PEDro de que estudos que revelem efeitos significativos do tratamento e que obtenham pontuação elevada na escala PEDro não fornecem, necessariamente, evidência de que o tratamento tenha sido clinicamente útil. Adicionalmente, importa saber se o efeito do tratamento foi suficientemente expressivo para poder ser considerado clinicamente justificável, se os efeitos positivos do mesmo suplantam os negativos, e também aferir a relação de custo-eficácia do tratamento. A escala não deve ser utilizada para comparar a "qualidade" de ensaios realizados em diferentes áreas de terapia, porque nalgumas áreas da prática da fisioterapia não é possível satisfazer todos os itens da escala." do original "The PEDro scale is based on the Delphi list developed by Verhagen and colleagues at the Department of Epidemiology, University of Maastricht (Verhagen AP et al (1998). The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of Clinical Epidemiology, 51(12):1235-41). The list is based on \"expert consensus\" not, for the most part, on empirical data. Two additional items not on the Delphi list (PEDro scale items 8 and 10) have been included in the PEDro scale. As more empirical data comes to hand it may become possible to \"weight\" scale items so that the PEDro score reflects the importance of individual scale items. The purpose of the PEDro scale is to help the users of the PEDro database rapidly identify which of the known or suspected randomised clinical trials (ie RCTs or CCTs) archived on the PEDro database are likely to be internally valid (criteria 2-9), and could have sufficient statistical information to make their results interpretable (criteria 10-11). An additional criterion (criterion 1) that relates to the external validity (or \"generalisability\" or \"applicability\" of the trial) has been retained so that the Delphi list is complete, but this criterion will not be used to calculate the PEDro score reported on the PEDro web site. The PEDro scale should not be used as a measure of the \"validity\" of a study's</p>	5	4/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
12 (cont.)	conclusions. In particular, we caution users of the PEDro scale that studies which show significant treatment effects and which score highly on the PEDro scale do not necessarily provide evidence that the treatment is clinically useful. Additional considerations include whether the treatment effect was big enough to be clinically worthwhile, whether the positive effects of the treatment outweigh its negative effects, and the cost-effectiveness of the treatment. The scale should not be used to compare the "quality" of trials performed in different areas of therapy, primarily because it is not possible to satisfy all scale items in some areas of physiotherapy practice."?	5	4/6
13	Concorda com a tradução da explicação geral dos critérios "Indicações para a administração da escala PEDro: Todos os critérios Só se atribui pontuação quando um critério foi inequivocamente satisfeito. Se numa leitura literal do relatório do ensaio for possível considerar que o critério não foi satisfeito, esse critério não deve receber pontuação." do original "Notes on administration of the PEDro scale: All criteria Points are only awarded when a criterion is clearly satisfied. If on a literal reading of the trial report it is possible that a criterion was not satisfied, a point should not be awarded for that criterion."?	5	3/6
14	Concorda com a tradução da explicação do critério 1 "Este critério pode considerar-se satisfeito quando o relatório descreve as origens dos sujeitos e a lista de critérios utilizados para determinar quais os sujeitos elegíveis para participar no estudo." do original "This criterion is satisfied if the report describes the source of subjects and a list of criteria used to determine who was eligible to participate in the study."?	4	1/6
15	Concorda com a tradução da explicação do critério 2 "Considera-se que num determinado estudo houve distribuição aleatória se o relatório refere que a distribuição dos sujeitos foi aleatória. O método de aleatoriedade não precisa de ser explicitado. Sorteios tais como lançamento de dados ou moeda ao ar devem considerar-se procedimentos de distribuição aleatória. Procedimentos de distribuição quase-aleatória tais como os que se efectuam a partir do número de registo no hospital, da data de nascimento, ou de alternância, não satisfazem este critério." do original "A study is considered to have used random allocation if the report states that allocation was random. The precise method of randomisation need not be specified. Procedures such as coin-tossing and dice-rolling should be considered random. Quasi-randomisation allocation procedures such as allocation by hospital record number or birth date, or alternation, do not satisfy this criterion."?	5	4/6
16	Concorda com a tradução da explicação do critério 3 "Distribuição sigilosa significa que a pessoa que determinou a elegibilidade do sujeito para participar no ensaio desconhecia, no momento da decisão, o grupo a que o sujeito iria pertencer. Deve atribuir-se um ponto a este critério, mesmo que não se diga que a distribuição foi sigilosa, quando o relatório refere que a distribuição foi feita a partir de envelopes fechados ou que a distribuição implicou o contacto com o responsável pela distribuição dos sujeitos por grupos, e este último não estava implicado no ensaio." do original "Concealed allocation means that the person who determined if a subject was eligible for inclusion in the trial was unaware, when this decision was made, of which group the subject would be allocated to. A point is awarded for this criteria, even if it is not stated that allocation was concealed, when the report states that allocation was by sealed opaque envelopes or that allocation involved contacting the holder of the allocation schedule who was "off-site"."	5	3/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
17	Concorda com a tradução da explicação do critério 4 "No mínimo, nos estudos de intervenções terapêuticas, o relatório deve descrever pelo menos uma medida do estado da condição a ser tratada e pelo menos uma (diferente) medida de resultado-chave que caracteriza o ponto de partida. O avaliador deve assegurar-se de que não é possível prever, com base nas condições de partida, diferenças de resultados para os diversos grupos. " do original "At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one (different) key outcome measure at baseline. The rater must be satisfied that the groups' outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented."?	4	0/6
18	Concorda com a tradução da explicação dos critérios 4,7-11 "Resultados-chave são resultados que fornecem o primeiro indicador da eficácia (ou falta de eficácia) da terapia. Na maioria dos estudos, há mais de uma variável que pode ser usada como indicador de eficácia." do original "Key outcomes are those outcomes which provide the primary measure of the effectiveness (or lack of effectiveness) of the therapy. In most studies, more than one variable is used as an outcome measure."?	5	4/6
19	Concorda com a tradução da explicação dos critérios 5-7 "Estudo Cego em relação ao sujeito, terapeuta ou avaliador significa que a pessoa em questão (sujeito, terapeuta ou avaliador) não conhece qual o grupo em que o sujeito é integrado. O estudo só é cego em relação aos sujeitos e aos terapeutas se for possível esperar-se que os mesmos sejam incapazes de distinguir entre os tratamentos aplicados aos diferentes grupos. Nos ensaios em que os resultados-chave são relatados pelo próprio (por exemplo, escala visual análoga, registo diário da dor), o estudo é considerado cego em relação ao avaliador se o for em relação ao sujeito. " do original "Blinding means the person in question (subject, therapist or assessor) did not know which group the subject had been allocated to. In addition, subjects and therapists are only considered to be "blind" if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In trials in which key outcomes are self-reported (eg, visual analogue scale, pain diary), the assessor is considered to be blind if the subject was blind."?	5	3/6
20	Concorda com a tradução da explicação do critério 8 "Este critério só se considera satisfeito se o relatório referir explicitamente tanto o número de sujeitos inicialmente integrados nos grupos como o número de sujeitos para os quais se obtiveram medidas de resultados-chave. Nos ensaios em que os resultados são medidos em diferentes momentos, um resultado-chave tem de ter sido medido em mais de 85% dos sujeitos num destes momentos." do original "This criterion is only satisfied if the report explicitly states both the number of subjects initially allocated to groups and the number of subjects from whom key outcome measures were obtained. In trials in which outcomes are measured at several points in time, a key outcome must have been measured in more than 85% of subjects at one of those points in time."?	5	3/6
21	Concorda com a tradução da explicação do critério 9 "Uma análise de intenção de tratamento significa que, quando os sujeitos não receberam tratamento (ou a condição de controlo) tal como previsto, e quando se encontram disponíveis medidas de resultados, a análise foi efectuada como se os sujeitos tivessem recebido o tratamento (ou a condição de controlo) que lhes tido sido atribuído inicialmente. Este critério é satisfeito, mesmo que não haja menção de análise por intenção de	3	2/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
21 (cont.)	tratamento, se o relatório referir explicitamente que todos os sujeitos receberam o tratamento ou condição de controlo previstos." do original "An intention to treat analysis means that, where subjects did not receive treatment (or the control condition) as allocated, and where measures of outcomes were available, the analysis was performed as if subjects received the treatment (or control condition) they were allocated to. This criterion is satisfied, even if there is no mention of analysis by intention to treat, if the report explicitly states that all subjects received treatment or control conditions as allocated."?	3	2/6
22	Concorda com a tradução da explicação do critério 10 "Uma comparação inter-grupal estatística implica uma comparação estatística de um grupo com outro. Conforme o desenho do estudo, isto pode implicar uma comparação de dois ou mais tratamentos, ou a comparação do tratamento com a condição de controlo. A análise pode ser uma simples comparação dos resultados medidos após a administração do tratamento, ou a comparação da mudança num grupo em relação à mudança no outro (quando se usou uma análise factorial de variância para analisar os dados, esta última é frequentemente descrita como interação grupo x tempo). A comparação pode assumir a forma de uma estimativa (por exemplo, a diferença média ou a diferença mediana, ou uma diferença nas proporções, ou um número necessário para tratar, ou um risco relativo ou um rácio de risco) e respectivo intervalo de confiança." do original "A between-group statistical comparison involves statistical comparison of one group with another. Depending on the design of the study, this may involve comparison of two or more treatments, or comparison of treatment with a control condition. The analysis may be a simple comparison of outcomes measured after the treatment was administered, or a comparison of the change in one group with the change in another (when a factorial analysis of variance has been used to analyse the data, the latter is often reported as a group ´ time interaction). The comparison may be in the form hypothesis testing (which provides a "p" value, describing the probability that the groups differed only by chance) or in the form of an estimate (for example, the mean or median difference, or a difference in proportions, or number needed to treat, or a relative risk or hazard ratio) and its confidence interval."?	3	2/6
23	Concorda com a tradução da explicação do critério 11 "Um ponto de medida é uma medida da dimensão do efeito do tratamento. O efeito do tratamento pode ser descrito como uma diferença nos resultados do grupo, ou como o resultado em todos os (ou em cada um dos) grupos. Medidas de variabilidade incluem desvios-padrão, erros-padrão, intervalos de confiança, amplitudes interquartis (ou outras amplitudes de quantis), e amplitudes de variação. Os pontos de medida e/ou as medidas de variabilidade podem ser apresentados graficamente (por exemplo, os DP podem ser apresentados como barras de erro numa figura) desde que aquilo que é representado seja inequivocamente identificável (por exemplo, desde que fique claro se as barras de erro representam DP ou EP). Quando os resultados são categóricos, considera-se que este critério foi satisfeito se o número de sujeitos em cada categoria é dada para cada grupo." do original "A point measure is a measure of the size of the treatment effect. The treatment effect may be described as a difference in group outcomes, or as the outcome in (each of) all groups. Measures of variability include standard deviations, standard errors, confidence intervals, interquartile ranges (or other quantile ranges), and ranges. Point measures and/or measures of variability may be provided graphically (for example, SDs may be given as error bars in a Figure) as long as it is clear	5	4/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
23 (cont.)	what is being graphed (for example, as long as it is clear whether error bars represent SDs or SEs). Where outcomes are categorical, this criterion is considered to have been met if the number of subjects in each category is given for each group."?	5	4/6

Ronda 2

Nº	Pergunta	Pontuação	
		Moda	Val 5
1 (antes 2)	Concorda com a tradução do item 2 da escala de PEDro "Os sujeitos foram aleatoriamente distribuídos por grupos (num estudo crossover, os sujeitos foram colocados em grupos de forma aleatória de acordo com o tratamento recebido)" do original "subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)." ?	5	5/6
2 (antes 3)	Concorda com a tradução do item 3 da escala de PEDro "A distribuição dos sujeitos foi cega" do original "allocation was concealed."?	5	5/6
3 (antes 4)	Concorda com a tradução do item 4 da escala de PEDro "Inicialmente, os grupos eram semelhantes no que diz respeito aos indicadores de prognóstico mais importantes." do original "the groups were similar at baseline regarding the most important prognostic indicators." ?	5	6/6
4 (antes 5)	Concorda com a tradução do item 5 da escala de PEDro "Todos os sujeitos participaram de forma cega no estudo" do original "there was blinding of all subjects."?	5	6/6
5 (antes 6)	Concorda com a tradução do item 6 da escala de PEDro "Todos os terapeutas que administraram a terapia fizeram-no de forma cega." do original "there was blinding of all therapists who administered the therapy."?	5	5/6
6 (antes 7)	Concorda com a tradução do item 7 da escala de PEDro "Todos os avaliadores que mediram pelo menos um resultado-chave, fizeram-no de forma cega." do original "there was blinding of all assessors who measured at least one key outcome."?	5	6/6
7 (antes 8)	Concorda com a tradução do item 8 da escala de PEDro "Medições de pelo menos um resultado-chave foram obtidas em mais de 85% dos sujeitos inicialmente distribuídos pelos grupos " do original "measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups."?	5	6/6
8 (antes 9)	Concorda com a tradução do item 9 da escala de PEDro "Todos os sujeitos a partir dos quais se apresentam medições de resultados receberam o tratamento ou a condição de controlo conforme a distribuição ou, quando não foi esse o caso, fez-se a análise dos dados para pelo menos um dos resultados-chave por "intenção de tratamento"?" do original "all subjects	5	5/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
8 (antes 9)	for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by 'intention to treat'."?	5	5/6
9 (antes 11)	Concorda com a tradução do item 11 da escala de PEDro "O estudo apresenta tanto medidas de localização como medidas de variabilidade para pelo menos um resultado-chave" do original "the study provides both point measures and measures of variability for at least one key outcome."?	5	4/6
10 (antes 12)	Concorda com a tradução da introdução da escala de PEDro "A escala PEDro baseia-se na lista de Delphi, desenvolvida por Verhagen e colegas no Departamento de Epidemiologia, da Universidade de Maastricht (Verhagen AP et al (1988). "The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus". Journal of Clinical Epidemiology, 51(12):1235-41). A lista, na sua maior parte, baseia-se num "consenso de peritos" e não em dados empíricos. Incluíram-se na escala de PEDro dois itens adicionais, que não constavam da lista de Delphi (os itens 8 e 10 da escala de PEDro). À medida que forem disponibilizados mais dados empíricos, pode vir a ser possível ponderar os itens da escala de forma a que a pontuação obtida a partir da aplicação da escala PEDro reflita a importância de cada um dos itens da escala. O objectivo da escala PEDro consiste em auxiliar os utilizadores da base de dados PEDro a identificar rapidamente quais dos ensaios clínicos randomizados, conhecidos ou suspeitos, (ou seja, ECR ou ECC) arquivados na base de dados PEDro poderão ter validade interna (critérios 2-9), e poderão conter suficiente informação estatística para que os seus resultados possam ser interpretados (critérios 10-11). Um critério adicional (critério 1) que diz respeito à validade externa (ou "potencial de generalização" ou "aplicabilidade" do ensaio) foi conservado para que a lista de Delphi esteja completa, mas este critério não será usado para calcular a pontuação PEDro apresentada no endereço PEDro na internet. A escala PEDro não deverá ser usada como uma medida da "validade" das conclusões de um estudo. Advertimos, muito especialmente, os utilizadores da escala PEDro de que estudos que revelem efeitos significativos do tratamento e que obtenham pontuação elevada na escala PEDro não fornecem, necessariamente, evidência de que o tratamento seja clinicamente útil. Adicionalmente, importa saber se o efeito do tratamento foi suficientemente expressivo para poder ser considerado clinicamente justificável, se os efeitos positivos superam os negativos, e aferir a relação de custo-eficácia do tratamento. A escala não deve ser utilizada para comparar a "qualidade" de ensaios realizados em diferentes áreas de terapia, porque nalgumas áreas da prática da fisioterapia não é possível satisfazer todos os itens da escala." do original "The PEDro scale is based on the Delphi list developed by Verhagen and colleagues at the Department of Epidemiology, University of Maastricht (Verhagen AP et al (1998). The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of Clinical Epidemiology, 51(12):1235-41). The list is based on "expert consensus" not, for the most part, on empirical data. Two additional items not on the Delphi list (PEDro scale items 8 and 10) have been included in the PEDro scale. As more empirical data comes to hand it may become possible to "weight" scale items so that the PEDro score reflects the importance of individual scale items. The purpose of the PEDro scale is to help the users of the PEDro database rapidly identify which of the known or suspected randomised clinical trials (ie RCTs or CCTs) archived on the PEDro database are likely to be internally valid	5	5/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
10 (antes 12) (cont.)	(criteria 2-9), and could have sufficient statistical information to make their results interpretable (criteria 10-11). An additional criterion (criterion 1) that relates to the external validity (or "generalisability" or "applicability" of the trial) has been retained so that the Delphi list is complete, but this criterion will not be used to calculate the PEDro score reported on the PEDro web site. The PEDro scale should not be used as a measure of the "validity" of a study's conclusions. In particular, we caution users of the PEDro scale that studies which show significant treatment effects and which score highly on the PEDro scale do not necessarily provide evidence that the treatment is clinically useful. Additional considerations include whether the treatment effect was big enough to be clinically worthwhile, whether the positive effects of the treatment outweigh its negative effects, and the cost-effectiveness of the treatment. The scale should not be used to compare the "quality" of trials performed in different areas of therapy, primarily because it is not possible to satisfy all scale items in some areas of physiotherapy practice."?	5	5/6
11 (antes 13)	Concorda com a tradução da explicação geral dos critérios "Indicações para a administração da escala PEDro: Todos os critérios A pontuação só será atribuída quando um critério for claramente satisfeito. Se numa leitura literal do relatório do ensaio existir a possibilidade de um critério não ter sido satisfeito, esse critério não deve receber pontuação." do original "Notes on administration of the PEDro scale: All criteria Points are only awarded when a criterion is clearly satisfied. If on a literal reading of the trial report it is possible that a criterion was not satisfied, a point should not be awarded for that criterion."?	5	5/6
12 (antes 14)	Concorda com a tradução da explicação do critério 1 "Este critério pode considerar-se satisfeito quando o relatório descreve a origem dos sujeitos e a lista de requisitos utilizados para determinar quais os sujeitos elegíveis para participar no estudo." do original "This criterion is satisfied if the report describes the source of subjects and a list of criteria used to determine who was eligible to participate in the study."?	5	5/6
13 (antes 15)	Concorda com a tradução da explicação do critério 2 "Considera-se que num determinado estudo houve distribuição aleatória se o relatório referir que a distribuição dos sujeitos foi aleatória. O método de aleatoriedade não precisa de ser explícito. Procedimentos tais como lançamento de dados ou moeda ao ar devem considerar-se de distribuição aleatória. Procedimentos de distribuição quase-aleatória tais como os que se efectuam a partir do número de registo hospitalar, da data de nascimento, ou de alternância, não satisfazem este critério." do original "A study is considered to have used random allocation if the report states that allocation was random. The precise method of randomisation need not be specified. Procedures such as coin-tossing and dice-rolling should be considered random. Quasi-randomisation allocation procedures such as allocation by hospital record number or birth date, or alternation, do not satisfy this criterion."?	5	6/6
14 (antes 16)	Concorda com a tradução da explicação do critério 3 "Distribuição cega significa que a pessoa que determinou a elegibilidade do sujeito para participar no ensaio desconhecia, quando a decisão foi tomada, o grupo a que o sujeito iria pertencer. Deve atribuir-se um ponto a este critério, mesmo que não se diga que a distribuição foi cega, quando o relatório refere que a distribuição foi feita a partir de envelopes opacos fechados ou que a distribuição implicou o contacto com o responsável pela distribuição dos sujeitos por grupos, e este último não estava implicado no ensaio." do original "Concealed allocation means that the person who	5	5/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
14 (antes 16) (cont.)	determined if a subject was eligible for inclusion in the trial was unaware, when this decision was made, of which group the subject would be allocated to. A point is awarded for this criteria, even if it is not stated that allocation was concealed, when the report states that allocation was by sealed opaque envelopes or that allocation involved contacting the holder of the allocation schedule who was “off-site”.?”	5	5/6
15 (antes 17)	Concorda com a tradução da explicação do critério 4 "No mínimo, nos estudos de intervenções terapêuticas, o relatório deve descrever pelo menos uma medida de severidade da condição a ser tratada e pelo menos uma (diferente) medida de resultado-chave que caracterize o ponto de partida. O avaliador deve assegurar-se de que não é possível prever, com base nas condições de partida, diferenças clinicamente significativas de resultados para os diversos grupos. Este critério é satisfeito mesmo que somente sejam apresentados os dados iniciais do estudo." do original "At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one (different) key outcome measure at baseline. The rater must be satisfied that the groups' outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented.”?	5	4/6
16 (antes 18)	Concorda com a tradução da explicação dos critérios 4,7-11 "Resultados-chave são resultados que fornecem o primeiro indicador da eficácia (ou falta de eficácia) da terapia. Na maioria dos estudos, há mais de uma variável que pode ser usada como medida de resultados." do original "Key outcomes are those outcomes which provide the primary measure of the effectiveness (or lack of effectiveness) of the therapy. In most studies, more than one variable is used as an outcome measure.”?	5	4/6
17 (antes 19)	Concorda com a tradução da explicação dos critérios 5-7 "Estar cego ao estudo significa que a pessoa em questão (sujeito, terapeuta ou avaliador) não conhece qual o grupo em que o sujeito é integrado. Mais ainda, sujeitos e terapeutas só são considerados “cegos” se for possível esperar-se que os mesmos sejam incapazes de distinguir entre os tratamentos aplicados aos diferentes grupos. Nos ensaios em que os resultados-chave são relatados pelo próprio (por exemplo, escala visual análoga, registo diário da dor), o avaliador é considerado “cego” se o sujeito foi “cego”." do original "Blinding means the person in question (subject, therapist or assessor) did not know which group the subject had been allocated to. In addition, subjects and therapists are only considered to be “blind” if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In trials in which key outcomes are self-reported (eg, visual analogue scale, pain diary), the assessor is considered to be blind if the subject was blind.”?	5	4/6
18 (antes 20)	Concorda com a tradução da explicação do critério 8 "Este critério só se considera satisfeito se o relatório referir explicitamente tanto o número de sujeitos inicialmente integrados nos grupos como o número de sujeitos a partir dos quais se obtiveram medidas de resultados-chave. Nos ensaios em que os resultados são medidos em diferentes momentos no tempo, um resultado-chave tem de ter sido medido em mais de 85% dos sujeitos num destes momentos." do original "This criterion is only satisfied if the report explicitly states both the number of subjects initially allocated to groups and the number of subjects from whom key outcome measures were obtained. In trials in which outcomes are measured at several points in time, a key outcome must have been measured in more than 85% of subjects at one of those points in time.”?	5	6/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
19 (antes 21)	Concorda com a tradução da explicação do critério 9 "Uma análise de intenção de tratamento significa que, quando os sujeitos não receberam tratamento (ou a condição de controlo) conforme o grupo atribuído, e quando se encontram disponíveis medidas de resultados, a análise foi efectuada como se os sujeitos tivessem recebido o tratamento (ou a condição de controlo) que lhes tido sido atribuído inicialmente. Este critério é satisfeito, mesmo que não haja menção de análise por intenção de tratamento, se o relatório referir explicitamente que todos os sujeitos receberam o tratamento ou condição de controlo, conforme a distribuição por grupos." do original "An intention to treat analysis means that, where subjects did not receive treatment (or the control condition) as allocated, and where measures of outcomes were available, the analysis was performed as if subjects received the treatment (or control condition) they were allocated to. This criterion is satisfied, even if there is no mention of analysis by intention to treat, if the report explicitly states that all subjects received treatment or control conditions as allocated."?	5	5/6
20 (antes 22)	Concorda com a tradução da explicação do critério 10 "Uma comparação estatística inter-grupos implica uma comparação estatística de um grupo com outro. Conforme o desenho do estudo, isto pode implicar uma comparação de dois ou mais tratamentos, ou a comparação do tratamento com a condição de controlo. A análise pode ser uma simples comparação dos resultados medidos após a administração do tratamento, ou a comparação das alterações num grupo em relação às alterações no outro (quando se usou uma análise factorial de variância para analisar os dados, esta última é frequentemente descrita como interacção grupo x tempo). A comparação pode apresentar-se sob a forma de hipóteses (através de um valor de p, descrevendo a probabilidade dos grupos diferirem apenas por acaso) ou assumir a forma de uma estimativa (por exemplo, a diferença média ou a diferença mediana, ou uma diferença nas proporções, ou um número necessário para tratar, ou um risco relativo ou um rácio de risco) e respectivo intervalo de confiança." do original "A between-group statistical comparison involves statistical comparison of one group with another. Depending on the design of the study, this may involve comparison of two or more treatments, or comparison of treatment with a control condition. The analysis may be a simple comparison of outcomes measured after the treatment was administered, or a comparison of the change in one group with the change in another (when a factorial analysis of variance has been used to analyse the data, the latter is often reported as a group x time interaction). The comparison may be in the form hypothesis testing (which provides a "p" value, describing the probability that the groups differed only by chance) or in the form of an estimate (for example, the mean or median difference, or a difference in proportions, or number needed to treat, or a relative risk or hazard ratio) and its confidence interval."?	5	5/6
21 (antes 23)	Concorda com a tradução da explicação do critério 11 "Uma Medida de localização é uma medida da dimensão do efeito do tratamento. O efeito do tratamento pode ser descrito como uma diferença nos resultados do grupo, ou como o resultado em todos os (ou em cada um dos) grupos. Medidas de variabilidade incluem desvios-padrão, erros-padrão, intervalos de confiança, amplitudes interquartis (ou outras amplitudes de quantis), e amplitudes de variação. As medidas de localização e/ou as medidas de variabilidade podem ser apresentadas graficamente (por exemplo, os DP's podem ser apresentados como barras de erro numa figura) desde que aquilo que é representado seja inequivocamente identificável (por exemplo, desde que fique claro se as barras de erro	4	3/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
21 (antes 23) (cont.)	representam DP's ou EP's). Quando os resultados são categóricos, considera-se que este critério foi cumprido se o número de sujeitos em cada categoria é dada para cada grupo." do original "A point measure is a measure of the size of the treatment effect. The treatment effect may be described as a difference in group outcomes, or as the outcome in (each of) all groups. Measures of variability include standard deviations, standard errors, confidence intervals, interquartile ranges (or other quantile ranges), and ranges. Point measures and/or measures of variability may be provided graphically (for example, SDs may be given as error bars in a Figure) as long as it is clear what is being graphed (for example, as long as it is clear whether error bars represent SDs or SEs). Where outcomes are categorical, this criterion is considered to have been met if the number of subjects in each category is given for each group."?	4	3/6

Ronda 3

Nº	Pergunta	Pontuação	
		Moda	Val 5
1 (antes 11)	Concorda com a tradução do item 11 da escala de Pedro "O estudo apresenta tanto medidas de precisão como medidas de variabilidade para pelo menos um resultado-chave" do original "the study provides both point measures and measures of variability for at least one key outcome."?	5	6/6
2 (antes 17)	Concorda com a tradução da explicação do critério 4 "No mínimo, nos estudos de intervenções terapêuticas, o relatório deve descrever pelo menos uma medida de gravidade da condição a ser tratada e pelo menos uma (diferente) medida de resultado-chave que caracterize o ponto de partida. O avaliador deve assegurar-se de que, com base nas condições de partida, não é possível prever diferenças clinicamente significativas de resultados para os diversos grupos. Este critério é satisfeito mesmo que somente sejam apresentados os dados iniciais do estudo." do original "At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one (different) key outcome measure at baseline. The rater must be satisfied that the groups' outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented."?	5	3/6
3 (antes 18)	Concorda com a tradução da explicação dos critérios 4,7-11 "Resultados-chave são resultados que fornecem o indicador primário da eficácia (ou falta de eficácia) da terapia. Na maioria dos estudos, utilizam mais do que uma variável como medida de resultados." do original "Key outcomes are those outcomes which provide the primary measure of the effectiveness (or lack of effectiveness) of the therapy. In most studies, more than one variable is used as an outcome measure."?	5	5/6

Nº	Pergunta	Pontuação	
		Moda	Val 5
4 (antes 19)	Concorda com a tradução da explicação dos critérios 5-7 "Ser cego para o estudo significa que a pessoa em questão (sujeito, terapeuta ou avaliador) não conhece qual o grupo em que o sujeito é integrado. Mais ainda, sujeitos e terapeutas só são considerados "cegos" se for possível esperar-se que os mesmos sejam incapazes de distinguir entre os tratamentos aplicados aos diferentes grupos. Nos ensaios em que os resultados-chave são relatados pelo próprio (por exemplo, escala visual análoga, registo diário da dor), o avaliador é considerado "cego" se o sujeito foi "cego"." do original "Blinding means the person in question (subject, therapist or assessor) did not know which group the subject had been allocated to. In addition, subjects and therapists are only considered to be "blind" if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In trials in which key outcomes are self-reported (eg, visual analogue scale, pain diary), the assessor is considered to be blind if the subject was blind."?	5	5/6
5 (antes 23)	Concorda com a tradução da explicação do critério 11 "Uma Medida de precisão é uma medida da dimensão do efeito do tratamento. O efeito do tratamento pode ser descrito como uma diferença nos resultados do grupo, ou como o resultado em todos os (ou em cada um dos) grupos. Medidas de variabilidade incluem desvios-padrão(DP's), erros-padrão (EP's), intervalos de confiança, amplitudes interquartis (ou outras amplitudes de quantis), e amplitudes de variação. As medidas de localização e/ou as medidas de variabilidade podem ser apresentadas graficamente (por exemplo, os DP's podem ser apresentados como barras de erro numa figura) desde que aquilo que é representado seja inequivocamente identificável (por exemplo, desde que fique claro se as barras de erro representam DP's ou EP's). Quando os resultados são relativos a variáveis categóricas, considera-se que este critério foi cumprido se o número de sujeitos em cada categoria é dado para cada grupo." do original "A point measure is a measure of the size of the treatment effect. The treatment effect may be described as a difference in group outcomes, or as the outcome in (each of) all groups. Measures of variability include standard deviations, standard errors, confidence intervals, interquartile ranges (or other quantile ranges), and ranges. Point measures and/or measures of variability may be provided graphically (for example, SDs may be given as error bars in a Figure) as long as it is clear what is being graphed (for example, as long as it is clear whether error bars represent SDs or SEs). Where outcomes are categorical, this criterion is considered to have been met if the number of subjects in each category is given for each group."?	5	5/6

Ronda 4

Nº	Pergunta	Pontuação	
		Moda	Val 5
1 (antes 17)	Concorda com a tradução da explicação do critério 4 "No mínimo, nos estudos de intervenções terapêuticas, o relatório deve descrever pelo menos uma medida da gravidade da condição a ser tratada e pelo menos uma (diferente) medida de resultado-chave que caracterize o ponto de partida. O avaliador deve assegurar-se de que, com base nas condições de prognóstico de início, não seja possível prever diferenças clinicamente significativas dos resultados, para os diversos grupos. Este critério é atingido mesmo que somente sejam apresentados os dados iniciais do estudo." do original "At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one (different) key outcome measure at baseline. The rater must be satisfied that the groups' outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented."?	5	6/6

Apêndice V – Tradução PEDro Scale de Inglês para Português versão 2

Escala de PEDro

Critério	Sim	Não	Onde:
1. Os critérios de elegibilidade foram especificados?			
2. Os sujeitos foram aleatoriamente distribuídos por grupos (num estudo crossover, os sujeitos foram colocados em grupos de forma aleatória de acordo com o tratamento recebido).			
3. A distribuição dos sujeitos foi cega.			
4. Inicialmente, os grupos eram semelhantes no que diz respeito aos indicadores de prognóstico mais importantes.			
5. Todos os sujeitos participaram de forma cega no estudo.			
6. Todos os fisioterapeutas que administraram a terapia fizeram-no de forma cega.			
7. Todos os avaliadores que mediram pelo menos um resultado-chave, fizeram-no de forma cega.			
8. Medições de pelo menos um resultado-chave foram obtidas em mais de 85% dos sujeitos inicialmente distribuídos pelos grupos.			
9. Todos os sujeitos a partir dos quais se apresentaram medições de resultados receberam o tratamento ou a condição de controlo conforme a distribuição ou, quando não foi esse o caso, fez-se a análise dos dados para pelo menos um dos resultados-chave por “intenção de tratamento”.			
10. Os resultados das comparações estatísticas inter-grupos foram descritos para pelo menos um resultado-chave.			
11. O estudo apresenta tanto medidas de precisão como medidas de variabilidade para pelo menos um resultado-chave.			
Pontuação			

A escala PEDro baseia-se na lista de Delphi, desenvolvida por Verhagen e colegas no Departamento de Epidemiologia, da Universidade de Maastricht (Verhagen AP et al (1988). “The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus”. *Journal of Clinical Epidemiology*, 51(12):1235-41). A lista, na sua maior parte,

baseia-se num “consenso de peritos” e não em dados empíricos. Incluíram-se na escala de PEDro dois itens adicionais, que não constavam da lista de Delphi (os itens 8 e 10 da escala de PEDro). À medida que forem disponibilizados mais dados empíricos, pode vir a ser possível ponderar os itens da escala de forma a que a pontuação obtida a partir da aplicação da escala PEDro reflecta a importância de cada um dos itens da escala.

O objectivo da escala PEDro consiste em auxiliar os utilizadores da base de dados PEDro a identificar rapidamente quais dos ensaios clínicos randomizados, conhecidos ou suspeitos, (ou seja, ECR ou ECC) arquivados na base de dados PEDro poderão ter validade interna (critérios 2-9), e poderão conter suficiente informação estatística para que os seus resultados possam ser interpretados (critérios 10-11). Um critério adicional (critério 1) que diz respeito à validade externa (ou “potencial de generalização” ou “aplicabilidade” do ensaio) foi mantido para que a *Delphi list* esteja completa, mas este critério não será usado para calcular a pontuação PEDro apresentada no endereço PEDro na internet.

A escala PEDro não deverá ser usada como uma medida da “validade” das conclusões de um estudo. Advertimos, muito especialmente, os utilizadores da escala PEDro de que estudos que revelem efeitos significativos do tratamento e que obtenham pontuação elevada na escala PEDro não fornecem, necessariamente, evidência de que o tratamento seja clinicamente útil. Adicionalmente, importa saber se o efeito do tratamento foi suficientemente expressivo para poder ser considerado clinicamente justificável, se os efeitos positivos superam os negativos, e aferir a relação de custo-eficácia do tratamento. A escala não deve ser utilizada para comparar a “qualidade” de ensaios realizados em diferentes áreas de terapia, principalmente porque nalgumas áreas da prática da fisioterapia não é possível satisfazer todos os itens da escala.

Indicações para a administração da escala PEDro

Todos os

critérios

A pontuação só será atribuída quando um critério for claramente satisfeito. Se numa leitura literal do relatório do ensaio existir a possibilidade de um critério não ter sido satisfeito, esse critério não deve receber pontuação.

Critério 1

Este critério pode considerar-se satisfeito quando o relatório descreve a origem dos sujeitos e a lista de requisitos utilizados para determinar quais os sujeitos elegíveis para participar no estudo.

Critério 2

Considera-se que num determinado estudo houve distribuição aleatória se o relatório referir que a distribuição dos sujeitos foi aleatória. O método de aleatoriedade não precisa de ser explícito. Procedimentos tais como lançamento de dados ou moeda ao ar devem considerar-se de distribuição aleatória. Procedimentos de

- distribuição quase-aleatória tais como os que se efectuam a partir do número de registo hospitalar, da data de nascimento, ou de alternância, não satisfazem este critério.
- Critério 3 *Distribuição cega* significa que a pessoa que determinou a elegibilidade do sujeito para participar no ensaio desconhecia, quando a decisão foi tomada, o grupo a que o sujeito iria pertencer. Deve atribuir-se um ponto a este critério, mesmo que não se diga que a distribuição foi cega, quando o relatório refere que a distribuição foi feita a partir de envelopes opacos fechados ou que a distribuição implicou o contacto com o responsável pela distribuição dos sujeitos por grupos, e este último não estava implicado no ensaio.
- Critério 4 No mínimo, nos estudos de intervenções terapêuticas, o relatório deve descrever pelo menos uma medida da gravidade da condição a ser tratada e pelo menos uma (diferente) medida de resultado-chave que caracterize o ponto de partida. O examinador deve assegurar-se de que, com base nas condições de prognóstico de início, não seja possível prever diferenças clinicamente significativas dos resultados, para os diversos grupos. Este critério é atingido mesmo que somente sejam apresentados os dados iniciais do estudo.
- Critérios 4, 7-11 *Resultados-chave* são resultados que fornecem o indicador primário da eficácia (ou falta de eficácia) da terapia. Na maioria dos estudos, utilizam mais do que uma variável como medida de resultados.
- Critérios 5-7 *Ser cego para o estudo* significa que a pessoa em questão (sujeito, terapeuta ou avaliador) não conhece qual o grupo em que o sujeito é integrado. Mais ainda, sujeitos e terapeutas só são considerados “cegos” se for possível esperar-se que os mesmos sejam incapazes de distinguir entre os tratamentos aplicados aos diferentes grupos. Nos ensaios em que os resultados-chave são relatados pelo próprio (por exemplo, escala visual análoga, registo diário da dor), o avaliador é considerado “cego” se o sujeito foi “cego”.
- Critério 8 Este critério só se considera satisfeito se o relatório referir explicitamente tanto o número de sujeitos inicialmente integrados nos grupos como o número de sujeitos a partir dos quais se obtiveram medidas de resultados-chave. Nos ensaios em que os resultados são medidos em diferentes momentos no tempo, um resultado-chave tem de ter sido medido em mais de 85% dos sujeitos num destes momentos.
- Critério 9 Uma análise de *intenção de tratamento* significa que, quando os sujeitos não receberam tratamento (ou a condição de controlo)

conforme o grupo atribuído, e quando se encontram disponíveis medidas de resultados, a análise foi efectuada como se os sujeitos tivessem recebido o tratamento (ou a condição de controlo) que lhes tido sido atribuído inicialmente. Este critério é satisfeito, mesmo que não seja referida a análise por intenção de tratamento, se o relatório referir explicitamente que todos os sujeitos receberam o tratamento ou condição de controlo, conforme a distribuição por grupos.

Critério 10

Uma *comparação estatística inter-grupos* implica uma comparação estatística de um grupo com outro. Conforme o desenho do estudo, isto pode implicar uma comparação de dois ou mais tratamentos, ou a comparação do tratamento com a condição de controlo. A análise pode ser uma simples comparação dos resultados medidos após a administração do tratamento, ou a comparação das alterações num grupo em relação às alterações no outro (quando se usou uma análise factorial de variância para analisar os dados, esta última é frequentemente descrita como interacção grupo x tempo). A comparação pode apresentar-se sob a forma de hipóteses (através de um valor de p, descrevendo a probabilidade dos grupos diferirem apenas por acaso) ou assumir a forma de uma estimativa (por exemplo, a diferença média ou a diferença mediana, ou uma diferença nas proporções, ou um número necessário para tratar, ou um risco relativo ou um rácio de risco) e respectivo intervalo de confiança.

Critério 11

Uma *Medida de precisão* é uma medida da dimensão do efeito do tratamento. O efeito do tratamento pode ser descrito como uma diferença nos resultados do grupo, ou como o resultado em todos os (ou em cada um dos) grupos. Medidas de variabilidade incluem desvios-padrão (DP's), erros-padrão (EP's), intervalos de confiança, amplitudes interquartis (ou outras amplitudes de quantis), e amplitudes de variação. As medidas de precisão e/ou as medidas de variabilidade podem ser apresentadas graficamente (por exemplo, os DP's podem ser apresentados como barras de erro numa figura) desde que aquilo que é representado seja inequivocamente identificável (por exemplo, desde que fique claro se as barras de erro representam DP's ou EP's). Quando os resultados são relativos a variáveis categóricas, considera-se que este critério foi cumprido se o número de sujeitos em cada categoria é dado para cada grupo.

Apêndice VI – Retro-tradução

PEDro Scale

<i>Criterion</i>	Yes	No	Where:
12. Were the eligibility criteria specified?			
13. Subjects were randomly allocated amongst groups (in a crossover study, subjects were randomly put into groups in accordance with the treatment received).			
14. Allocation of the subjects was concealed.			
15. Initially, the groups were similar as regards the most important prognosis indicators.			
16. All the subjects participated in the study in blind fashion.			
17. All the physiotherapists who administered therapy did so in blind fashion.			
18. All the assessors who measured at least one key result did so in blind fashion.			
19. Measures of at least one key result were obtained in over 85% of the subjects initially allocated amongst the groups.			
20. All the subjects for whom measures of results were provided received the treatment or the control condition in accordance with the allocation or, where then this was not the case, an analysis of the data was carried out for at least one of the key results by "treatment intention".			
21. The results of the inter-group statistical comparisons were provided for at least one key result.			
22. The study provides both accuracy measures as well as variability measures for at least one key result.			
	Scoring		

The PEDro scale is based on the Delphi list developed by Verhagen and colleagues at the Epidemiology Department of the University of Maastricht (Verhagen AP et al (1988). "The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus". *Journal of Clinical Epidemiology*, 51(12):1235-41). The list is based, in the main, on "expert consensus" and not on empirical data. Two additional items were included on the PEDro scale which were not to be found on the Delphi list (items 8 and 10 of the PEDro scale). As and when more empirical data is provided, it may

become possible to score the items of the scale in such a way that the score obtained from the PEDro scale application reflects the importance of each of the items on the scale.

The aim of the PEDro scale is to aid users of the PEDro data base to rapidly identify which of the known or suspected randomised, clinical trials, (in other words, RCTs or CCTs) filed on the PEDro data base may have internal validity (criteria 2-9) and may contain enough statistical information so that its results can be interpreted (criteria 10-11). An additional criterion (criterion 1) which relates to the external validity (or “generalisability” or “applicability” of the trial) was maintained so that the *Delphi list* is complete, but this criterion shall not be used to calculate the PEDro score presented at the PEDro website.

The PEDro scale should not be used as a measure of the “validity” of the conclusions of a study. We would particularly like to warn the users of the PEDro scale that studies which reveal major treatment effects and which obtain high scoring on the PEDro scale do not necessarily provide proof that the treatment is clinically useful. In addition, it is important to know whether the treatment effect has been sufficiently expressive to be able to be considered clinically justifiable, whether the positive effects outweigh the negative ones, and to ascertain the cost-effectiveness of the treatment. The scale should not be used to compare the “quality” of the trials carried out in different therapy areas, mainly because in some fields of physiotherapy practice, it is not possible to satisfy all the items on the scale.

Indications for the administration of the PEDro scale

All criteria	<u>Points shall only be assigned when a criterion has clearly been satisfied.</u> If, when reading the trial report in literal fashion, there is the possibility that a criterion has not been satisfied, no points shall be assigned to said criterion.
Criterion 1	This criterion may be regarded as having been satisfied when the report describes the origin of the subjects and the list of requirements used to determine which subjects are eligible to participate in the study.
Criterion 2	There shall be deemed to have been random allocation in a given study if the report states that the allocation of the subjects was random. The randomisation method does not need to be precisely specified. Procedures such as the rolling of dice or tossing a coin in the air should be regarded as being of random allocation. Quasi-random allocation procedures such as by hospital record number, date of birth or alternation do not meet this criterion.
Criterion 3	<i>Concealed allocation</i> means that the person who determined the eligibility of the subject to participate in the trial was unaware,

- when the decision was taken, to which group the subject would belong. One point shall be assigned to this criterion, even if it has not been said that the allocation was concealed, when the report says that the allocation has been carried out by way of sealed, opaque envelopes or that the allocation entailed contact with the party responsible for the allocation of the subjects amongst groups and the latter had not been involved in the trial.
- Criterion 4 At the very least, in therapeutic intervention studies, the report shall describe at least one measure of the seriousness of the condition to be treated and at least one (different) key result measure which characterises the point of departure. The examiner should make sure that, based on the starting prognosis conditions, it is not possible to foresee clinically significant differences from the results for the various groups. This criterion is achieved even if only the initial study data is presented.
- Criteria 4, 7-11 *Key results* are results that provide the primary effectiveness (or ineffectiveness) indicator of the therapy. In the majority of the studies, more than one variable is used to measure results.
- Criteria 5-7 *Study blinding* means that the person in question (subject, therapist or assessor) is unaware which group the subject belongs to. Furthermore, subjects and therapists are only regarded as “blind” if it is possible to expect that they would be unable to distinguish between the treatments applied to the various groups. In the trials in which the key results are reported by the subject itself (for instance, visual analogue scale, pain diary), the assessor is regarded as “blind” if the subject was “blind”.
- Criterion 8 This criterion is only considered to have been met if the report explicitly states both the number of subjects initially integrated in the groups as well as the number of subjects based on whom measures of key results have been obtained. In trials in which the results are measured at different moments in time, a key result must have been measured in over 85% of the subjects at one of these moments.
- Criterion 9 A *treatment intention* analysis means that when the subjects did not receive treatment (or the control condition) depending on the group allocated and where measures of results are available, the analysis has been carried out as if the subjects had received the treatment (or the control condition) which had been allocated to them initially. This criterion is satisfied, even if the treatment intention analysis is not referred to, if the report explicitly states that all the subjects received the treatment or control condition, depending on the allocation by groups.

- Criterion 10 An *intergroup statistical comparison* entails a statistical comparison between one group with another. Depending on how the study has been designed, this may entail a comparison between two or more treatments, or the comparison of the treatment with the control condition. The analysis may be a simple comparison of the results measured after the administration of the treatment, or the comparison of the alterations in one group with the alterations in the other (when factorial analysis of variance has been used to analyse the data, the latter is frequently described as a group x time interaction). The comparison may take the form of hypotheses (by way of a value of p , describing the likelihood of the groups differing merely by chance) or of an estimate (for example, the mean difference or the median difference, or a difference in proportions, or a number needed for treatment, or a relative risk or a risk ratio) and the respective confidence interval.
- Criterion 11 An *Accuracy Measurement* is a measure of the size of the treatment effect. The treatment effect may be described as a difference in the group results or as the result in all (or in each of) the groups. Variability measures include standard deviations (SD's), standard errors (SE's), confidence intervals, interquartile ranges (or other quantile ranges), and variation ranges. The accuracy measures and/or variability measures may be presented graphically (for instance, the SD's may be presented as error bars in a figure) provided that what is being identifiable is unequivocally identifiable (for instance, provided that it is clear whether the error bars represent SD's or SE's). When the results relate to categorical variables, it is considered that this criterion has been met if the number of subjects in each category is provided for each group.
- .

Apêndice VII – Consentimento Informado Pré-teste

Consentimento Informado

Título: Tradução e adaptação transcultural da PEDro Scale para a língua e cultura portuguesa: um instrumento de avaliação de ensaios clínicos

Autora: Cláudia Costa, aluna do Mestrado em Ciências da Fisioterapia na Faculdade de Motricidade Humana (FMH)

Orientador: Professor Doutor Jan Maria Cabri

A prática baseada na evidência tem-se tornado uma das maiores preocupações hoje em dia, quer para os profissionais de saúde quer para os utentes que usufruem da prestação de cuidados. O facto de a investigação estar em ascensão faz com que por vezes não seja produzida da melhor forma, torna-se então necessária a existência de um instrumento de medida que permita verificar e avaliar o rigor da metodologia científica.

O presente estudo tem como objectivo proceder à tradução e adaptação cultural da *PEDro scale*, uma escala que permite avaliar parâmetros relativos à qualidade da metodologia de ensaios clínicos em Fisioterapia.

Face a este objectivo será necessário que preencha um questionário de caracterização sócio-demográfica e avalie 3 artigos com a versão portuguesa da referida escala. Este questionário é anónimo e confidencial, sem quaisquer tipos de desvantagens para o participante, servindo apenas para fins estatísticos.

Pretende-se que a informação acima esteja clara, caso compreenda tudo o que lhe foi apresentado, esteja livre de dúvidas e queira participar por favor assine em baixo.

Assinatura do Participante

_____ Data: ____/____/____

Desde já obrigada,

A Investigadora

_____ Data ____/____/____

Apêndice VIII – Formulário Pré-teste

Formulário de compreensão**PEDro Scale**

Data de nascimento: __/__/__

Sexo: _____

Habilitações Literárias: _____

Anos de profissão: _____

Data preenchimento: __/__/__

Artigo 1

Início preenchimento: __h__m

Fim preenchimento: __h__m

Tp de preenchimento: __h__m

Artigo 2

Início preenchimento: __h__m

Fim preenchimento: __h__m

Tp de preenchimento: __h__m

Artigo 3

Início preenchimento: __h__m

Fim preenchimento: __h__m

Tp de preenchimento: __h__m

Escala de PEDro

Os itens a seguir referidos suscitaram dúvidas?

1. Os critérios de elegibilidade foram especificados?

1 ☐
Incompreensível2 ☐
Muitas
dúvidas3 ☐
Algumas
dúvidas4 ☐
Compreensível5 ☐
Claro

2. Os sujeitos foram aleatoriamente distribuídos por grupos (num estudo crossover, os sujeitos foram colocados em grupos de forma aleatória de acordo com o tratamento recebido).

1 ☐
Incompreensível2 ☐
Muitas
dúvidas3 ☐
Algumas
dúvidas4 ☐
Compreensível5 ☐
Claro

3. A distribuição dos sujeitos foi cega.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

4. Inicialmente, os grupos eram semelhantes no que diz respeito aos indicadores de prognóstico mais importantes.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

5. Todos os sujeitos participaram de forma cega no estudo.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

6. Todos os fisioterapeutas que administraram a terapia fizeram-no de forma cega.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

7. Todos os avaliadores que mediram pelo menos um resultado-chave, fizeram-no de forma cega.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

8. Medições de pelo menos um resultado-chave foram obtidas em mais de 85% dos sujeitos inicialmente distribuídos pelos grupos.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

9. Todos os sujeitos a partir dos quais se apresentaram medições de resultados receberam o tratamento ou a condição de controlo conforme a distribuição ou, quando não foi esse o caso, fez-se a análise dos dados para pelo menos um dos resultados-chave por “intenção de tratamento”.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

10. Os resultados das comparações estatísticas inter-grupos foram descritos para pelo menos um resultado-chave.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

11. O estudo apresenta tanto medidas de precisão como medidas de variabilidade para pelo menos um resultado-chave.

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

Instruções Escala de PEDro

As instruções foram claras ou sentiu alguma dificuldade na compreensão das mesmas?

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

Se sim o que não foi explicito?

Existem algumas expressões ou palavras que não tenha compreendido? Se sim indique quais.

Existe algo em falta nas instruções?

Opinião geral

No seu geral o instrumento foi claro quanto ao seu objectivo e utilidade?

1 ☐
Incompreensível

2 ☐
Muitas
dúvidas

3 ☐
Algumas
dúvidas

4 ☐
Compreensível

5 ☐
Claro

Criticas e sugestões

Obrigado!

Apêndice IX – Escala PEDro (versão portuguesa)

Escala PEDRO (Portuguese version)

<i>Critério</i>	Sim	Não	Onde:
1. Os critérios de elegibilidade foram especificados?			
2. Os sujeitos foram aleatoriamente distribuídos por grupos (num estudo crossover, os sujeitos foram colocados em grupos de forma aleatória de acordo com o tratamento recebido).			
3. A distribuição dos sujeitos foi cega.			
4. Inicialmente, os grupos eram semelhantes no que diz respeito aos indicadores de prognóstico mais importantes.			
5. Todos os sujeitos participaram de forma cega no estudo.			
6. Todos os fisioterapeutas que administraram a terapia fizeram-no de forma cega.			
7. Todos os avaliadores que mediram pelo menos um resultado-chave, fizeram-no de forma cega.			
8. Medições de pelo menos um resultado-chave foram obtidas em mais de 85% dos sujeitos inicialmente distribuídos pelos grupos.			
9. Todos os sujeitos a partir dos quais se apresentaram medições de resultados receberam o tratamento ou a condição de controlo conforme a distribuição ou, quando não foi esse o caso, fez-se a análise dos dados para pelo menos um dos resultados-chave por “intenção de tratamento”.			
10. Os resultados das comparações estatísticas inter-grupos foram descritos para pelo menos um resultado-chave.			
11. O estudo apresenta tanto medidas de precisão como medidas de variabilidade para pelo menos um resultado-chave.			
Pontuação			

Introdução

A escala PEDro baseia-se na lista de Delphi, desenvolvida por Verhagen e colegas no Departamento de Epidemiologia, da Universidade de Maastricht (Verhagen AP et al (1988). “The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus”. *Journal of Clinical Epidemiology*, 51(12):1235-41). A lista, na sua maior parte,

baseia-se num “consenso de peritos” e não em dados empíricos. Incluíram-se na escala de PEDro dois itens adicionais, que não constavam da lista de Delphi (os itens 8 e 10 da escala de PEDro). À medida que forem disponibilizados mais dados empíricos, pode vir a ser possível ponderar os itens da escala de forma a que a pontuação obtida a partir da aplicação da escala PEDro reflecta a importância de cada um dos itens da escala.

O objectivo da escala PEDro consiste em auxiliar os utilizadores da base de dados PEDro a identificar rapidamente quais dos ensaios clínicos randomizados, conhecidos ou suspeitos, (ou seja, ECR ou ECC) arquivados na base de dados PEDro poderão ter validade interna (critérios 2-9), e poderão conter suficiente informação estatística para que os seus resultados possam ser interpretados (critérios 10-11). Um critério adicional (critério 1) que diz respeito à validade externa (ou “potencial de generalização” ou “aplicabilidade” do ensaio) foi mantido para que a *Delphi list* esteja completa, mas este critério não será usado para calcular a pontuação PEDro apresentada no endereço PEDro na internet.

A escala PEDro não deverá ser usada como uma medida da “validade” das conclusões de um estudo. Advertimos, muito especialmente, os utilizadores da escala PEDro de que estudos que revelem efeitos significativos do tratamento e que obtenham pontuação elevada na escala PEDro não fornecem, necessariamente, evidência de que o tratamento seja clinicamente útil. Adicionalmente, importa saber se o efeito do tratamento foi suficientemente expressivo para poder ser considerado clinicamente justificável, se os efeitos positivos superam os negativos, e aferir a relação de custo-eficácia do tratamento. A escala não deve ser utilizada para comparar a “qualidade” de ensaios realizados em diferentes áreas de terapia, principalmente porque nalgumas áreas da prática da fisioterapia não é possível satisfazer todos os itens da escala.

Indicações para a administração da escala PEDro

Todos os

critérios

A pontuação só será atribuída quando um critério for claramente satisfeito. Se numa leitura literal do relatório do ensaio existir a possibilidade de um critério não ter sido satisfeito, esse critério não deve receber pontuação.

Critério 1

Este critério pode considerar-se satisfeito quando o relatório descreve a origem dos sujeitos e a lista de requisitos utilizados para determinar quais os sujeitos elegíveis para participar no estudo.

Critério 2

Considera-se que num determinado estudo houve distribuição aleatória se o relatório referir que a distribuição dos sujeitos foi aleatória. O método de aleatoriedade não precisa de ser explícito. Procedimentos tais como lançamento de dados ou moeda ao ar devem considerar-se de distribuição aleatória. Procedimentos de

- distribuição quase-aleatória tais como os que se efectuam a partir do número de registo hospitalar, da data de nascimento, ou de alternância, não satisfazem este critério.
- Critério 3 *Distribuição cega* significa que a pessoa que determinou a elegibilidade do sujeito para participar no ensaio desconhecia, quando a decisão foi tomada, o grupo a que o sujeito iria pertencer. Deve atribuir-se um ponto a este critério, mesmo que não se diga que a distribuição foi cega, quando o relatório refere que a distribuição foi feita a partir de envelopes opacos fechados ou que a distribuição implicou o contacto com o responsável pela distribuição dos sujeitos por grupos, e este último não estava implicado no ensaio.
- Critério 4 No mínimo, nos estudos de intervenções terapêuticas, o relatório deve descrever pelo menos uma medida da gravidade da condição a ser tratada e pelo menos uma (diferente) medida de resultado-chave que caracterize o ponto de partida. O examinador deve assegurar-se de que, com base nas condições de prognóstico de início, não seja possível prever diferenças clinicamente significativas dos resultados, para os diversos grupos. Este critério é atingido mesmo que somente sejam apresentados os dados iniciais do estudo.
- Critérios 4, 7-11 *Resultados-chave* são resultados que fornecem o indicador primário da eficácia (ou falta de eficácia) da terapia. Na maioria dos estudos, utilizam mais do que uma variável como medida de resultados.
- Critérios 5-7 *Ser cego para o estudo* significa que a pessoa em questão (sujeito, terapeuta ou avaliador) não conhece qual o grupo em que o sujeito é integrado. Mais ainda, sujeitos e terapeutas só são considerados “cegos” se for possível esperar-se que os mesmos sejam incapazes de distinguir entre os tratamentos aplicados aos diferentes grupos. Nos ensaios em que os resultados-chave são relatados pelo próprio (por exemplo, escala visual análoga, registo diário da dor), o avaliador é considerado “cego” se o sujeito foi “cego”.
- Critério 8 Este critério só se considera satisfeito se o relatório referir explicitamente tanto o número de sujeitos inicialmente integrados nos grupos como o número de sujeitos a partir dos quais se obtiveram medidas de resultados-chave. Nos ensaios em que os resultados são medidos em diferentes momentos no tempo, um resultado-chave tem de ter sido medido em mais de 85% dos sujeitos num destes momentos.

- Critério 9 Uma análise de *intenção de tratamento* significa que, quando os sujeitos não receberam tratamento (ou a condição de controlo) conforme o grupo atribuído, e quando se encontram disponíveis medidas de resultados, a análise foi efectuada como se os sujeitos tivessem recebido o tratamento (ou a condição de controlo) que lhes tinha sido atribuído inicialmente. Este critério é satisfeito, mesmo que não seja referida a análise por intenção de tratamento, se o relatório referir explicitamente que todos os sujeitos receberam o tratamento ou condição de controlo, conforme a distribuição por grupos.
- Critério 10 Uma *comparação estatística inter-grupos* implica uma comparação estatística de um grupo com outro. Conforme o desenho do estudo, isto pode implicar uma comparação de dois ou mais tratamentos, ou a comparação do tratamento com a condição de controlo. A análise pode ser uma simples comparação dos resultados medidos após a administração do tratamento, ou a comparação das alterações num grupo em relação às alterações no outro (quando se usou uma análise factorial de variância para analisar os dados, esta última é frequentemente descrita como interacção grupo x tempo). A comparação pode apresentar-se sob a forma de hipóteses (através de um valor de p, descrevendo a probabilidade dos grupos diferirem apenas por acaso) ou assumir a forma de uma estimativa (por exemplo, a diferença média ou a diferença mediana, ou uma diferença nas proporções, ou um número necessário para tratar, ou um risco relativo ou um rácio de risco) e respectivo intervalo de confiança.
- Critério 11 Uma *Medida de precisão* é uma medida da dimensão do efeito do tratamento. O efeito do tratamento pode ser descrito como uma diferença nos resultados do grupo, ou como o resultado em todos os (ou em cada um dos) grupos. Medidas de variabilidade incluem desvios-padrão(DP's), erros-padrão (EP's), intervalos de confiança, amplitudes interquartis (ou outras amplitudes de quantis), e amplitudes de variação. As medidas de precisão e/ou as medidas de variabilidade podem ser apresentadas graficamente (por exemplo, os DP's podem ser apresentados como barras de erro numa figura) desde que aquilo que é representado seja inequivocamente identificável (por exemplo, desde que fique claro se as barras de erro representam DP's ou EP's). Quando os resultados são relativos a variáveis categóricas, considera-se que este critério foi cumprido se o número de sujeitos em cada categoria é dado para cada grupo.

Apêndice X – Relatório Final

Translation and cultural adaptation of PEDro scale into Portuguese (Portugal) language: Escala de PEDro

Methods

Translation and cultural adaptation

The Portuguese translation followed the sequential forward and backward translation approach according to Beaton et al (2000) and Wild et al (2005):

- 1) Preparation – Authors authorization and involvement. Preparation of the methodology;
- 2) Forward Translation - translation of PEDro scale original language version into Portuguese (Portugal) by translator 1 (a bilingual Portuguese native-speaker);
- 3) Reconciliation - comparing and merging more than one expert opinion of the forward translation (first version). This was done by a web based Delphi Consensus Process: a panel of 6 experts evaluated and criticised the first translation. The evaluation was done through an ordinary scale (1 – do not agree to 5 – totally agree with the translation) and had to be consensual (5 or more experts' ratings of 5 points). Then the investigator performed the corrections and proceeded to another round until all the questions were consensual. After 4 rounds all questions were consensual about semantic equivalency. The Delphi process produced the second version of the forward translation.
- 4) Back Translation - translation of the Portuguese version back into the original language by translator 2 (a bilingual English native speaker);
- 5) Back Translation Review and Report- comparison of the back-translated version with the original to highlight and investigate discrepancies between the original and the reconciled translation by the authors of the PEDro Scale. The back-translated version was semantically equivalent to the PEDro Scale original version (report done by the author Anne Moseley and Leonardo Costa);
- 6) Pre-test Cognitive Debriefing - testing the instrument on a small group in order to test alternative wording and to check understanding, interpretation, and cultural relevance of the translation. This second version was applied by 20 physiotherapists for 3 published RCTs (2 in Portuguese and 1 in English - appendix 1). The 3 RCTs reports were part of the PEDro database. The agreement level was 75%. A comprehension survey with an ordinal scale (1 - unclear to 5 - clear) was filled by them also, conferring the understandable level (4);
- 7) Review of Cognitive Debriefing Results and Finalization - comparison of the participants' interpretation of the second version to highlight and

amend discrepancies, and composition of the third and final Portuguese version;

- 8) Proofreading - final review of the translation to highlight and correct any typographic, grammatical or other errors;
- 9) Final version and Report – the final version of Escala de PEDro report written at the end of the process documenting the development of each translation.

References

1. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25(24):3186-91.
2. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8(2):94-104.

Appendix 1

RCT report 1 – Azevedo, D. C., De Carvalho, S. C., Leal, E. S., Damasceno, S. P. & Ferreira, M. L. (Influence of the range of motion (ROM) limitation on the shoulder flexibility improvement after a six-week training) [Portuguese] *Revista Brasileira de Medicina do Esporte* 2008 Mar-Apr;14(2):119-121

RCT report 2 – Knobel, R., Radunz, V. & Carraro, T. E. (Use of transcutaneous electric nerve stimulation for pain relief in labour: a possible way to care the child-bearer) [Portuguese] *Texto & Contexto-Enfermagem* 2005;14(2):229-236

RCT report 3 – Ashburn, A., Fazakarley, L., Ballinger, C., Pickering, R., McLellan, L. D. & Fitton, C. A. Randomised controlled trial of a home based exercise programme to reduce the risk of falling among people with Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2007;78(7):678-84

Anexos

Anexo I – PEDro Scale (versão original)

PEDro Scale

1.	eligibility criteria were specified	no <input type="checkbox"/> yes <input type="checkbox"/> where:
2.	subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)	no <input type="checkbox"/> yes <input type="checkbox"/> where:
3.	allocation was concealed	no <input type="checkbox"/> yes <input type="checkbox"/> where:
4.	the groups were similar at baseline regarding the most important prognostic indicators	no <input type="checkbox"/> yes <input type="checkbox"/> where:
5.	there was blinding of all subjects	no <input type="checkbox"/> yes <input type="checkbox"/> where:
6.	there was blinding of all therapists who administered the therapy	no <input type="checkbox"/> yes <input type="checkbox"/> where:
7.	there was blinding of all assessors who measured at least one key outcome	no <input type="checkbox"/> yes <input type="checkbox"/> where:
8.	measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups	no <input type="checkbox"/> yes <input type="checkbox"/> where:
9.	all subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by "intention to treat"	no <input type="checkbox"/> yes <input type="checkbox"/> where:
10.	the results of between-group statistical comparisons are reported for at least one key outcome	no <input type="checkbox"/> yes <input type="checkbox"/> where:
11.	the study provides both point measures and measures of variability for at least one key outcome	no <input type="checkbox"/> yes <input type="checkbox"/> where:

The PEDro scale is based on the Delphi list developed by Verhagen and colleagues at the Department of Epidemiology, University of Maastricht (*Verhagen AP et al (1998). The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of Clinical Epidemiology, 51(12):1235-41*). The list is based on "expert consensus" not, for the most part, on empirical data. Two additional items not on the Delphi list (PEDro scale items 8 and 10) have been included in the PEDro scale. As more empirical data comes to hand it may become possible to "weight" scale items so that the PEDro score reflects the importance of individual scale items.

The purpose of the PEDro scale is to help the users of the PEDro database rapidly identify which of the known or suspected randomised clinical trials (ie RCTs or CCTs) archived on the PEDro database are likely to be internally valid (criteria 2-9), and could have sufficient statistical information to make their results interpretable (criteria 10-11). An additional criterion (criterion 1) that relates to the external validity (or "generalisability" or "applicability" of the trial) has been retained so that the Delphi list is complete, but this criterion will not be used to calculate the PEDro score reported on the PEDro web site.

The PEDro scale should not be used as a measure of the "validity" of a study's conclusions. In particular, we caution users of the PEDro scale that studies which show significant treatment effects and which score highly on the PEDro scale do not necessarily provide evidence that the treatment is clinically useful. Additional

considerations include whether the treatment effect was big enough to be clinically worthwhile, whether the positive effects of the treatment outweigh its negative effects, and the cost-effectiveness of the treatment. The scale should not be used to compare the "quality" of trials performed in different areas of therapy, primarily because it is not possible to satisfy all scale items in some areas of physiotherapy practice.

Notes on administration of the PEDro scale:

- | | |
|------------------|--|
| All criteria | <p><u>Points are only awarded when a criterion is clearly satisfied.</u></p> <p>If on a literal reading of the trial report it is possible that a criterion was not satisfied, a point should not be awarded for that criterion.</p> |
| Criterion 1 | This criterion is satisfied if the report describes the source of subjects and a list of criteria used to determine who was eligible to participate in the study. |
| Criterion 2 | A study is considered to have used random allocation if the report states that allocation was random. The precise method of randomisation need not be specified. Procedures such as coin-tossing and dice-rolling should be considered random. Quasi-randomisation allocation procedures such as allocation by hospital record number or birth date, or alternation, do not satisfy this criterion. |
| Criterion 3 | <i>Concealed allocation</i> means that the person who determined if a subject was eligible for inclusion in the trial was unaware, when this decision was made, of which group the subject would be allocated to. A point is awarded for this criteria, even if it is not stated that allocation was concealed, when the report states that allocation was by sealed opaque envelopes or that allocation involved contacting the holder of the allocation schedule who was "off-site". |
| Criterion 4 | At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one (different) key outcome measure at baseline. The rater must be satisfied that the groups' outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented. |
| Criteria 4, 7-11 | <i>Key outcomes</i> are those outcomes which provide the primary measure of the effectiveness (or lack of effectiveness) of the therapy. In most studies, more than one variable is used as an outcome measure. |

- Criterion 5-7 *Blinding* means the person in question (subject, therapist or assessor) did not know which group the subject had been allocated to. In addition, subjects and therapists are only considered to be “blind” if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In trials in which key outcomes are self-reported (eg, visual analogue scale, pain diary), the assessor is considered to be blind if the subject was blind.
- Criterion 8 This criterion is only satisfied if the report explicitly states *both* the number of subjects initially allocated to groups *and* the number of subjects from whom key outcome measures were obtained. In trials in which outcomes are measured at several points in time, a key outcome must have been measured in more than 85% of subjects at one of those points in time.
- Criterion 9 An *intention to treat* analysis means that, where subjects did not receive treatment (or the control condition) as allocated, and where measures of outcomes were available, the analysis was performed as if subjects received the treatment (or control condition) they were allocated to. This criterion is satisfied, even if there is no mention of analysis by intention to treat, if the report explicitly states that all subjects received treatment or control conditions as allocated.
- Criterion 10 A *between-group* statistical comparison involves statistical comparison of one group with another. Depending on the design of the study, this may involve comparison of two or more treatments, or comparison of treatment with a control condition. The analysis may be a simple comparison of outcomes measured after the treatment was administered, or a comparison of the change in one group with the change in another (when a factorial analysis of variance has been used to analyse the data, the latter is often reported as a group \times time interaction). The comparison may be in the form hypothesis testing (which provides a “p” value, describing the probability that the groups differed only by chance) or in the form of an estimate (for example, the mean or median difference, or a difference in proportions, or number needed to treat, or a relative risk or hazard ratio) and its confidence interval.
- Criterion 11 A *point measure* is a measure of the size of the treatment effect. The treatment effect may be described as a difference in group outcomes, or as the outcome in (each of) all groups. *Measures of variability* include standard deviations, standard errors, confidence intervals, interquartile ranges (or other quantile ranges), and ranges. Point measures and/or measures of variability may be provided graphically (for example, SDs may

be given as error bars in a Figure) as long as it is clear what is being graphed (for example, as long as it is clear whether error bars represent SDs or SEs). Where outcomes are categorical, this criterion is considered to have been met if the number of subjects in each

Anexo II – Autorização Autores PEDro Scale

A autorização foi dada via e-mail:

Hello Jan and Claudia

I have now heard back from the rest of the PEDro team and have a unanimous vote to proceed with collaborating on the translation of the scale into Portuguese.

I would be happy to comment on a draft protocol or have further discussions to develop the plan for reliability testing when I get back.

Regards

Anne Moseley

***Anexo III – Relatório Retrotradução vs Versão Original
PEDro Scale***

O relatório foi enviado via e-mail:

Hi Jan and Claudia,

I have carefully read your back translation and will comment on it.

Firstly, it is not expected that the back translation will match 100% with the original English version; this is due to the language's diversity, synonyms and cultural aspects from each language. So I am not surprised to read some different words there. We have to bear in mind that Portuguese is a quite rich language having much more words than English.

For example the term "intention-to-treat" is extremely scientific, technical and relatively new. I don't believe that people without good evidence-based medicine training would be able to translate and back translate 100% accurately. The same thing is applicable for "between-group comparisons", "point measures", "baseline" etc...

I think what is important was that you guys found Portuguese words that are equivalent to the ones written on the original PEDro scale. As one of the oldest raters of PEDro and a Portuguese speaker I can say with confidence that your steps were performed correctly and the scale is grammatically equivalence with the English version.

Having saying that I can say that Claudia can finish her thesis!

However, the next question is: Is the Portuguese version of PEDro scale retains the original clinimetric properties from the English version? We will know that after our study!

I hope this helps!

Have a great weekend!

Leonardo Costa
Research Fellow - PEDro
PhD Candidate – Musculoskeletal Division
The George Institute for International Health

Address: PO Box M201 Missenden Rd Sydney NSW 2050 Australia

Phone: 61 2 8238 2440

Fax: 61 2 9657 0301

Email: lcosta@george.org.au

Web: www.thegeorgeinstitute.org <<http://www.thegeorgeinstitute.org>>